

	APPLICANT:	Kronmal, Gregory S.	
	APPLICANT:	Lauer, Peter M.	
	APPLICANT:	Ruddy, David A.	
	APPLICANT:	Thomas, Winston	
	APPLICANT:	Tsuchihashi, Zenta	
	APPLICANT:	Wolff, Roger K.	
	TITLE OF INVENTION:	Megabase Transcript Map: No. 5872237el	
	TITLE OF INVENTION:	Sequences and Antibodies Thereto	
	NUMBER OF SEQUENCES:	31	
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE:	TOWNSEND and TOWNSEND and CREW LLP	
	STREET:	Two Embarcadero Center, 8th Floor	
	CITY:	San Francisco	
	STATE:	CA	
	COUNTRY:	USA	
	ZIP:	94111-3834	
	COMPUTER READABLE FORM:		
	MEDIUM TYPE:	Floppy disk	
	OPERATING SYSTEM:	IBM PC compatible	
	SOFTWARE:	PatentIn Release #1.0, Version #1.30	
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER:	US/08/724,394A	
	FILING DATE:	01-OCT-1996	
	CLASSIFICATION:	536	
	ATTORNEY/AGENT INFORMATION:		
	NAME:	Fitts, Renee A.	
	REGISTRATION NUMBER:	35,136	
	REFERENCE/DOCKET NUMBER:	017957-000100	
	TELEPHONE:	415-576-0200	
	TELEFAX:	415-576-0300	
	INFORMATION FOR SEQ ID NO:	21:	
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	246240 base pairs	
	TYPE:	nucleic acid	
	STRANDEDNESS:	not relevant	
	TOPOLOGY:	not relevant	
	MOLECULE TYPE:	cDNA	
	FEATURE:		
	NAME/KEY:	misc.feature	
	LOCATION:	1..246240	
	OTHER INFORMATION:	/note= "HLA-H.CONTIG"	
	US-08-724-394A-20		
	Query Match	3.6%; Score 199.2; DB 2; Length 246240;	
	Best Local Similarity	61.7%; Pred.No. 2.9e-32;	
	Matches 439; Conservative	0; Mismatches 233; Indels 40; Gaps 6;	
QY	3199	TTTTAGTATTCATTTTATGTTTTTGATGTTTTTAAAAAGACAATCTCACCC	3258
DB	3311	TTTTTATAAGACTTTTTTTTTTTTTTTTTTTTTTTTGGACGGAGTCTGCTCT	3252
QY	3259	GTTACCAGCGGTGAGTGCAGTGGTGCATCATAGCTTCTTGAGTCTTGA	3318
DB	3251	GTGCCCAGGCTGGAGTGCAGTGGCGGGATCTCGGCTCACTCGGCTCGGCTCGGG	3192
QY	3319	CTCAAGCAATCCCTCCCTGCGCTGGCCCTCCAAAGTGTGGGA-TACAGT	3377
DB	3191	TTACGCCCATTCCTTGCCTTCAGCTTCCCAAGTAGCTGGGACTACAGGCGCGCCAC	3132
QY	3378	CATCTGGC--CTAGGATCCATTTAGATTAAAATATGATTTTAAATTTTAAATAATATG	3435
DB	3131	CGCCCGGCTAAATTTTGTATTTTGTAGTAGACAGCGGGITTCACCGTTT	3072
QY	3436	GCTAAATTTT-----ACCTTATGTAATGTGTAATCT	3466
DB	3071	GTCTCGATCTCTGACCTCGTGATCGCGCGCTCGGCTCCCAAAGTGTGGGATTACA	3012
QY	3467	GGTAATAATCTAGTTGTGCTTAAGTTTAAAGTCTTCCAAATAGCTTCATGTACG	3526
DB	3011	GGCGTAGGCAACGCGCGCGCTTATAAGACTTTTATAGATAAAACAGAATCTATATC	2952

; OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-21

	Query Match	3.6%	Score 199.2;	DB 2;	Length 246240;
	Best Local Similarity	61.7%;	Pred. No. 2.9e-32;		
	Matches 439; Conservative	0;	Mismatches 233;	Indels 40;	Gaps 6;
QY	3199	TTTTGTAGTATTCATTTAATGTTTTTTTATGTTTTTGATTTTTTAAAAAGACAATCTCACCCCT	3258		
Db	3311				
QY	3259	GTACCCAGCGCTGGAGTCAGTGGTGCAAATCATAGACTTCTGCAGCTCTTGAACCTCCTGGG	3318		
Db	3251	GTGCCCCAGCGTGGAGTCAGTGGCGGGATCTCGGCTCACTGCAAGCTCCGCCCTCCCGGG	3192		
QY	3319	CTCAAGCAATCTCTCGCCTTGGCCTCCCAAAGTGTGGGA-TACAGTCATGAGCCACTG	3377		
Db	3191	TTTACGCCATTCTCTGCTCAGCTCCCAAGTAGCTGGGACTACAGSGCCCCGCCACATA	3132		
QY	3378	CATCTGGC--CTAGGATCCATTAGATTAAAATATGCATTTAAATTTTAAATAAATATG	3435		
Db	3131	CGCCCGGCTAAATTTTGTGATTTTAGTAGAGACGGGTTTACCCTTTTAGCCGGGATG	3072		
QY	3436	GCTAAATTTT-----ACCTTATGTAATCTGTATACT	3466		
Db	3071	GTCTCGATCTCTGACCTCGTGATCGGCCCGCTCGGCCTCCCAAGTCTGGGATTACA	3012		
QY	3467	GGAATAAATCTAGTTTGTGCTTAAAGTTTAAAGTCTTTCCAATAAGCTTTCATGTACG	3526		
Db	3011	GGCGTGAGCCACCGCGCGCCCTTATAAAGACTTTTTAGATAAAAACAGATCTATATC	2952		
QY	3527	TGAGGGGAGACATTTTAAAGTGAACAGACAGCCAGGTGTGGTGCCTCAGCCCTGTAATCC	3586		
Db	2951	TCTAATGTATGTATTAAAGATGAACC-CGGCTGGGCGCAGTGGTCTACACCTGTAAATCC	2893		
QY	3587	CAGCACTCTGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTTGGAGTTTCAGACACAGCT	3646		
Db	2892	CAGCACTTTGGAGGCCGAGGCGGTGATCACTTCAGGTCAAGGAGTTTCAGACACAGCT	2833		
QY	3647	GAGCAACATGCANAAACCTGTTCCT-----ATAACAAAAATTAGCCGGGCATGGTGGC	3700		
Db	2832	GGCAAAACACGGCAAACCTCTCTCTACTAAAAATACAAAAATTAGCTGGCATGTGGTC	2773		
QY	3701	ATGTCCCTGTGGTCCCAGCTACTAGG-GGCTGAGGAGGAGAATCTTTGAGAGCCAGGA	3759		
Db	2772	AGGCACCTGTAAATCCAGCTGCTTTGGAGACTGAGGACGAGGAATCGTTTTTAAACCTGGGA	2713		
QY	3760	GGTCAGGCTGCCTGAGCAGTGTCTGCGGCACCTGCCTCCAGCTGGGTGCACAGGACCA	3819		
Db	2712	CGTGGGATTCAGTGAGCCAAAGCTGCACCACCTGCATCCAGCCTGGGCACAGGGTGA	2653		
QY	3820	GACCTTGCCTCAAAAAATAGAGAAAAATTTAAAAATAAATGGAACAACACT	3871		
Db	2652	GACACGCTCTCAAAAAATAAAAAATAAATAAATAAATAAATAAATAAATGACCCCTCT	2601		

RESULT 7

US-08-724-394A-22/c
; Sequence 22, Application US/08724394A

Patent No. 5872237

```

: GENERAL INFORMATION:
:
: APPLICANT: Feder, John N.
: APPLICANT: Kronmal, Gregory S.
: APPLICANT: Lauer, Peter M.
: APPLICANT: Ruddy, David A.
: APPLICANT: Thomas, Winston
: APPLICANT: Tsuchinashi, Zenta
: APPLICANT: Wolff, Roger K.
:
: TITLE OF INVENTION: Megabase Transcript Map: No. 58722372el
:
: NUMBER OF SEQUENCES: 31
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

```

STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/724,394A
 FILING DATE: 01-OCT-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Fitts, Renee A.
 REGISTRATION NUMBER: 35,136
 REFERENCE/DOCKET NUMBER: 017957-000100
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-576-0200
 TELEFAX: 415-576-0300
 INFORMATION FOR SEQ ID NO: 22:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 246240 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: cdna
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 1..246240
 OTHER INFORMATION: /note= "HLA-H.CONFIG"
 US-08-724-394A-22

[illegible]

Db 93229 CAGTCTCCCAAGTAGCTGAAATATACAGGTGTGCGCCACCATGCCAGGCTAATTTTGTAT 93288
QY 3397 TTAGATTAAATATGCAATTTTAAATTTTAAATATATATGCGCTAATTTTACCTTATGTA 3456
Db 93289 TTTTGTAGTGAGACAGGGGTTCACCATGGTGCCAGCGTATCTGCGAACTCCTGACCTCA 93348
QY 3457 TGTGTATACTGGTATATAAATCTAGTTTGTCTGCTGCTAAAGTTTAAAGTGCTTTCCCAATAAGC 3516
Db 93349 AGTG-ATCCACCTGTGTGGCCTTCCAAAGTCTGGAATTACAGGTGTGGCCACTGGC 93407
QY 3517 TT--CATGTAGTGAGGGGAGACACATTTAAAGTGAACAGACAG-CCAGGTGTGGTGGCTC 3573
Db 93408 CTGGCCAGTTTATATTTTAAACATTAATAAATGCTTCTGTGTCCAATCATGTGTGTCTC 93467
QY 3574 ACGCTGTAAATCCCAAGCACTCTGGGAGGCTGAGGTGAGGTGGGTGATCGCTTGAGCCCTGGAGT 3633
Db 93458 ACACCCATTAATCCAGCACTTTGGGAGGCACAGCTTGGAGAACTCACCTTGAGGCGCAGGAT 93527
QY 3634 TCAAGACGAGCTGAGCAACATGCGCAAAACCCCTGTTCTTATA-----ACAAAATTAGC 3687
Db 93528 TCGAGACGAGCTGGGCAACATGTTGAATCCCTGTCTCTACAAAAAATACAAAAATTAGC 93587
QY 3688 CGGCATGGTGCATGTCCTGTGTGCTCCAGCTACTAGG-GGCTGAGGCAGGAGATCT 3746
Db 93588 TGGCATGGTGGTGCATGCCCTGTAGTCCAGCTATTTGGGAGGCTGAGGTGGGAGATCA 93647
QY 3747 TTGGAGCCGAGAGGTCAAGGCTGACCTGACGAGTGTGGCGCCACTGCCTCCAGCCTG 3806
Db 93648 TTGGAGCCGAGAGTGTGGGCTGCGAGTGAGCTGAGATCAACCACTGACTCTCCAGGCTG 93707
QY 3807 GGTGACAGGACGACCTTGCTCAAAAAAATAAGAGAA 3846
Db 93708 GATGACAGCAAGGCCCGTTTCAAAAAAATAAGAGAA 93747

RESULT 14
US-09-754-250-3/c
; Sequence 3, Application US/09754250
; Patent No. 6376225
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001063
; CURRENT APPLICATION NUMBER: US/09/754,250
; CURRENT FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 111282
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(111282)
; OTHER INFORMATION: n = A,T,C or G
US-09-754-250-3

Query Match 3.4%; Score 190.2; DB 4; Length 111282;
Best Local Similarity 74.8%; Pred. No. 1.7e-30;
Matches 267; Conservative 0; Mismatches 83; Indels 7; Gaps 2;
QY 3542 AAAGTGAACACAGACAGCAGGTGGTGGCTGACGCTGTAATCCAGCACTCTGGGAGG 3601
Db 18630 AAAGTGAATGTTCAGGCCAGGTGGAGTGGCTGACGCTGTAATCCAGCACTTTGGGAGG 18571
QY 3602 CTGAGGTGGGTGATGCTGCTGAGCCCTGAGCTTCAAGACCAAGCTGTGAGCAACATGGCAAA 3661
Db 18570 TCAAGGCGGAGATCACTCTGAGGCGAGAGTTCGAGACCAAGCTGGCCAATATGGTCAA 18511
QY 3662 ACCCTGTTCT-----ATACAAAAATTAGCGGGCATGTGGCATGTGCTGTGGTCC 3715

Db 18510 ACCCGCTCTCTACCAAAAGTACAAAAATTAGCCGGCATGTGGCAGGTGCTGTGATCC 18451
QY 3716 CAGCTACTAGG-GCCTGAGCAGGAGAACTTTTGGAGCCCGCAGAGGTCAAGGCTGCACT 3774
Db 18450 CAGTACTTTGGAGGCTGAGCGCAGGAACAATCACTTGAACCCAGGAGGTGGAGGTGCACT 18391
QY 3775 GAGCAGTGTCTGGCCCACTGCACCTCCAGCCTGGGTGACAGGACGACGACCTTGGCCTCAAAA 3834
Db 18390 GAGACAAGATCGCACCACTGCACCTCCAGCCTGGCGACAGACAGCAAGACTGTCTCAAAAA 18331
QY 3835 AAATAAGAGAAGAAATTAATAATTAATGGAACAACACTACAAAAGAGCTGTTGCTCCTAG 3891
Db 18330 AAAAAAAGAGAAAAAAGAAAAAAGAAAAAAGTGTGTTACAG 18274

RESULT 15
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match 3.4%; Score 188.2; DB 4; Length 168575;
Best Local Similarity 64.7%; Pred. No. 5.2e-30;
Matches 314; Conservative 0; Mismatches 163; Indels 8; Gaps 2;
QY 3402 TTAATAATATGCATTTTAAATTTTAAATAATATATGCTAATTTTACCTTATGTAATGTGT 3461
Db 88918 TTAATAAATTTTAAATGTTCTGAGAGTGGATAGTGACGATGTTGCACTATAAAGTGAAT 88859
QY 3462 ATACTGGTAAATAATCTAGTTTGTGCTTAAAGTTTAAAGTGTCTTCCAATAAGCTTCAT 3521
Db 88858 ATACTTAATGTCTTGAACATATACACTTAAACACTGATTAAGTTGTGAATTTGTAATATAT 88799
QY 3522 GTAGCTGAGGGGAGACATTTTAAAGTGAACAGACAGCAGGCTGTGGTGGCTCAGCGCTGT 3581
Db 88798 GTATATTTTACCACAAATTTTAAATAATTAGTTGCCACTGGGAGTGGGCTCATGCCCTGT 88739
QY 3582 AATCCCAGCACTCTGGGAGGCTGAGGTGGGTGATCGCTTGGAGCCCTGGAGTTCAGAGCC 3641
Db 88738 AATCCCAGCACTTTGGGAGGCGGAGGTGGGTGGATCCCTCGATCAGATCAGGAGTTCGAGAC 88679
QY 3642 AGCCTGAGCAACATGGCAAAACCCCTGTTCT-----ATACAAAAATTAGCCGGCAT 3694
Db 88678 AGCTGGCCAACATAGTGAATCTCTCTACTAAAAAATACAAAAATTAGTGGGCTT 88619

QY	3695	GGTGGCATGTGCTGTGTGCCAGCTACT-AGGGGGCTGAGGCGAGGAATCTTTGGAGC	3753
Db	88618	GGTGGAGGAGGCTGTATCCAGCTACTCAGGAGGCTGAGGCGAGGAATCGCTTGAAC	88559
QY	3754	CCAGGAGTCAAGGCTGCAGTGCAGTGTGCGCCACTGCACTCCAGCCTGGGTGACA	3813
Db	88558	CCGGAGGTGGAGGCTGCAGTGCAGTGTGCGCCACTGCACTCCAGCCTGGGTGACA	88499
QY	3814	GGACCAGACCTTGGCTCAAAAAAATAAGAGAAATTAATAATGGAACAACACTAC	3873
Db	88498	GAGTGAGGCTCCCTCTCACAAAAAATAAGAGAAATTAATAATGGAACAACACTAC	88439
QY	3874	AAAGA	3878
Db	88438	ACAGA	88434

Search completed: June 19, 2003, 11:14:04
Job time : 239.577 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:06:39 ; Search time 739.811 seconds
(without alignments)
10937.110 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacatagtgctca.....tcaaaactctgacctcagag 5514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues
Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5512.8	100.0	5514	9	US-09-966-880A-9
2	4922.8	89.3	11204	9	US-09-966-880A-35
3	230.2	4.2	28818	10	US-09-764-877-2266
4	226.8	4.1	32204	9	US-09-764-872-517
5	226.8	4.1	32204	9	US-10-072-349-327
6	226.8	4.1	32204	10	US-09-764-855-327
7	225.4	4.1	28588	9	US-10-073-961-399
8	225.4	4.1	28588	10	US-09-764-887-399
9	221.6	4.0	10855	9	US-09-764-891-7965
10	221.2	4.0	32152	9	US-09-764-872-518
11	221.2	4.0	32152	9	US-10-072-349-328
12	221.2	4.0	32152	10	US-09-764-855-328
13	218.8	4.0	32134	9	US-09-764-891-6763
14	218.6	4.0	4421	9	US-10-216-373-6
15	215	3.9	19472	10	US-09-764-864-1698
16	214.4	3.9	17239	9	US-10-091-504-2326
17	214.4	3.9	17239	10	US-09-764-869-2326
18	213.6	3.9	32195	9	US-10-091-504-1605
19	213.6	3.9	32195	9	US-10-125-540-611

20	213.6	3.9	32195	9	US-10-125-540-617	Sequence 617, App
21	213.6	3.9	32195	10	US-09-764-870-611	Sequence 611, App
22	213.6	3.9	32195	10	US-09-764-870-617	Sequence 617, App
23	213.6	3.9	32195	10	US-09-764-869-1605	Sequence 1605, Ap
c 24	207.2	3.8	24132	9	US-10-074-095-661	Sequence 661, App
c 25	207.2	3.8	24132	10	US-09-764-860-661	Sequence 661, App
c 26	205.8	3.8	66686	10	US-09-736-960-86	Sequence 86, Appl
c 27	204.8	3.7	14769	9	US-09-764-891-7092	Sequence 7092, Ap
c 28	204.4	3.7	32195	9	US-10-091-504-1605	Sequence 1605, Ap
c 29	204.4	3.7	32195	9	US-10-125-540-611	Sequence 611, App
c 30	204.4	3.7	32195	9	US-10-125-540-617	Sequence 617, App
c 31	204.4	3.7	32195	10	US-09-764-870-611	Sequence 611, App
c 32	204.4	3.7	32195	10	US-09-764-870-617	Sequence 617, App
c 33	204.4	3.7	32195	10	US-09-764-869-1605	Sequence 1605, Ap
c 34	204.2	3.7	17996	10	US-09-764-877-2695	Sequence 2695, Ap
35	204	3.7	24132	9	US-10-074-095-661	Sequence 661, App
36	204	3.7	24132	10	US-09-764-860-661	Sequence 661, App
c 37	204	3.7	26657	10	US-09-810-673A-3	Sequence 3, Appl
c 38	202.8	3.7	32169	9	US-09-764-891-8605	Sequence 8605, Ap
c 39	202.8	3.7	52216	10	US-09-747-810-1	Sequence 1, Appl
40	201	3.6	7233	9	US-10-091-504-2123	Sequence 2123, Ap
41	201	3.6	7233	10	US-09-764-869-2123	Sequence 2123, Ap
42	200.6	3.6	23934	9	US-09-764-891-7210	Sequence 7210, Ap
c 43	200.6	3.6	23934	9	US-10-074-095-777	Sequence 777, App
c 44	200.6	3.6	23934	10	US-09-764-860-777	Sequence 777, App
45	200.6	3.6	23934	10	US-09-764-877-2536	Sequence 2536, Ap

ALIGNMENTS

RESULT 1
US-09-966-880A-9
; Sequence 9, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 5514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1)...(1031)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1032)...(1118)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1119)...(5514)
US-09-966-880A-9

Query Match 100.0%; Score 5512.8; DB 9; Length 5514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5514; Conservative 0; Mismatches 0; Gaps 0;

Oy 1 ACACGACGAATACATGGTCCAAAGCTAGGCTATTGTTTGAATAATCATCAAGGTATAGATG 60
|||||

Db 1 ACAGAGCAATACATGGTCCAGCTAGGGCTATTTGATTTGAAAAATCATCAAGGTATAGATG 60
QY 61 GTATCAAAAGGCTTGAGGCAGGAAGAGAGACAGAGACCCCTAGCTGCATTCCTTAGCATTTGCA 120
Db 61 GTATCAAAAGGCTTGAGGCAGGAAGAGAGACAGAGACCCCTAGCTGCATTCCTTAGCATTTGCA 120
QY 121 TCCCTAGCACCTGGCATAGTTTCCATTAACAGTAGGCGATGAAGTATCTPACTCAGTGAATA 180
Db 121 TCCCTAGCACCTGGCATAGTTTCCATTAACAGTAGGCGATGAAGTATCTPACTCAGTGAATA 180
QY 181 AATAGAATGCATATGGGCTCAGATAGGAGAGAGAAATAAATCTTTAATAGACCAAGTTC 240
Db 181 AATAGAATGCATATGGGCTCAGATAGGAGAGAGAAATAAATCTTTAATAGACCAAGTTC 240
QY 241 TATGAGAGACAAAATTAAGTCTTTTATTTGAAGATCTTAGCCCTGTTTTCCTCAAAATTCAG 300
Db 241 TATGAGAGACAAAATTAAGTCTTTTATTTGAAGATCTTAGCCCTGTTTTCCTCAAAATTCAG 300
QY 301 TGCAGCCAGTTAGACACTGATTCGTCTGGTGAACAAGCAATTTTGTATTTTGGGGAC 360
Db 301 TGCAGCCAGTTAGACACTGATTCGTCTGGTGAACAAGCAATTTTGTATTTTGGGGAC 360
QY 361 TGCCTGCTCTCTGACTCCAAATTAAGGATTTTCTTAAAGGATGGCTCATG 420
Db 361 TGCCTGCTCTCTGACTCCAAATTAAGGATTTTCTTAAAGGATGGCTCATG 420
QY 421 CAAAATCACCTTTTGGTGTAAATATCTAGTCTTCAAGCAATCTTGTAAATGCAATCAGA 480
Db 421 CAAAATCACCTTTTGGTGTAAATATCTAGTCTTCAAGCAATCTTGTAAATGCAATCAGA 480
QY 481 AAGAAAAAATCCATGGTTTGGGAGGCAAAATTTTGTGTCTAAATCTATATAACTGA 540
Db 481 AAGAAAAAATCCATGGTTTGGGAGGCAAAATTTTGTGTCTAAATCTATATAACTGA 540
QY 541 GTTCATTTTGTCTTAAGTCAAGAGCAGGAGCTGCTAGTGCCTGTCTGTACTGAGGTTTCAGAG 600
Db 541 GTTCATTTTGTCTTAAGTCAAGAGCAGGAGCTGCTAGTGCCTGTCTGTACTGAGGTTTCAGAG 600
QY 601 AGACTGTGGGAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACACAAATTAACCCAA 660
Db 601 AGACTGTGGGAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACACAAATTAACCCAA 660
QY 661 GAAGCTATTTAAATGCTCTTTAAGGTATTTACATAAAATPATTACATTTCTCATTTGCTGTT 720
Db 661 GAAGCTATTTAAATGCTCTTTAAGGTATTTACATAAAATPATTACATTTCTCATTTGCTGTT 720
QY 721 TATTTTGTGTATCATGATTTAATTTGAAGTGTCTACTGTTTACTGCCCTCCTGATCTTTGC 780
Db 721 TATTTTGTGTATCATGATTTAATTTGAAGTGTCTACTGTTTACTGCCCTCCTGATCTTTGC 780
QY 781 TAGCTATGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCCAAAGGAACCTTAAACATTTAA 840
Db 781 TAGCTATGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCCAAAGGAACCTTAAACATTTAA 840
QY 841 AGCAGAGCTGCCCTCAATGGTTTAAACCTGTGTGACTGTGCTATGACAGCCCCACCCACC 900
Db 841 AGCAGAGCTGCCCTCAATGGTTTAAACCTGTGTGACTGTGCTATGACAGCCCCACCCACC 900
QY 901 CATCTTCACTGATCCAAATCAGAGCAAGGCCGTTGGGTPACCTGGTGGGGGTGATGCT 960
Db 901 CATCTTCACTGATCCAAATCAGAGCAAGGCCGTTGGGTPACCTGGTGGGGGTGATGCT 960
QY 961 GTCAGGGGAGGAGGCCAAAAGGCAAGCTCAAAATTTGAATGTGAAGGGCCAAATGCACTGT 1020
Db 961 GTCAGGGGAGGAGGCCAAAAGGCAAGCTCAAAATTTGAATGTGAAGGGCCAAATGCACTGT 1020
QY 1021 CAGACTGAGACAGAGAACCATCATTTAATTTGAAGTGTGAGATTTTCTGGCCTGAGACTTGCA 1080
Db 1021 CAGACTGAGACAGAGAACCATCATTTAATTTGAAGTGTGAGATTTTCTGGCCTGAGACTTGCA 1080
QY 1081 GGGAGGCAAGAGACACTCTCGACACCACCTATGACAGGTAAAGAGGCAGTCTTCTCGTG 1140
Db 1081 GGGAGGCAAGAGACACTCTCGACACCACCTATGACAGGTAAAGAGGCAGTCTTCTCGTG 1140

QY 1141 GGTGATTTGCACCTGGCCTTCCCTCTCAGAGCAAAATCTGAGTAATGAGACTGGTAGCTATCCC 1200
Db 1141 GGTGATTTGCACCTGGCCTTCCCTCTCAGAGCAAAATCTGAGTAATGAGACTGGTAGCTATCCC 1200
QY 1201 TTTCTCTCATGTAACACTGTCTGACTGATAAGATCAAGTCTGATGTAATATGATATATATTTT 1260
Db 1201 TTTCTCTCATGTAACACTGTCTGACTGATAAGATCAAGTCTGATGTAATATGATATATATTTT 1260
QY 1261 TTGATCTGCTCTCCCTTTCTTCTTATTCAGATCTTATACGCTGTACGCCAAATCTTCTCTGT 1320
Db 1261 TTGATCTGCTCTCCCTTTCTTCTTATTCAGATCTTATACGCTGTACGCCAAATCTTCTCTGT 1320
QY 1321 TTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTGGCTACAATGT 1380
Db 1321 TTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTGGCTACAATGT 1380
QY 1381 ACTGATTCGCTCGAGATTTGTACCATGTTGAACTAAATTTATGTTAAATATATTAACA 1440
Db 1381 ACTGATTCGCTCGAGATTTGTACCATGTTGAACTAAATTTATGTTAAATATATTAACA 1440
QY 1441 TAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGGTAAATAGCAGTACTCTACTAAACAGG 1500
Db 1441 TAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGGTAAATAGCAGTACTCTACTAAACAGG 1500
QY 1501 TAGTCTTAATTTTTCGTAATAATTTTGTAAATATTCAACAGTAAACAACTTTGAAGACACA 1560
Db 1501 TAGTCTTAATTTTTCGTAATAATTTTGTAAATATTCAACAGTAAACAACTTTGAAGACACA 1560
QY 1561 CTTTCTTAGGAGGCGTTTACTGAAATATTTAGCTATAGTAAGAAAAATTTGTAATTTTGTAG 1620
Db 1561 CTTTCTTAGGAGGCGTTTACTGAAATATTTAGCTATAGTAAGAAAAATTTGTAATTTTGTAG 1620
QY 1621 AAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATTGTCTCCATTTATAGG 1680
Db 1621 AAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATTGTCTCCATTTATAGG 1680
QY 1681 AGACAAATTTCAATTCAGCAAGTTATTTAAATGTTTAAAGGCCCAATTTGTTAGCAGTTAATG 1740
Db 1681 AGACAAATTTCAATTCAGCAAGTTATTTAAATGTTTAAAGGCCCAATTTGTTAGCAGTTAATG 1740
QY 1741 GCACCTTTTACTATTAACATACTTTTCCATTTGTCAGACGTAGCTTAACTTTACCTCTTTAG 1800
Db 1741 GCACCTTTTACTATTAACATACTTTTCCATTTGTCAGACGTAGCTTAACTTTACCTCTTTAG 1800
QY 1801 GTGTGAATTTGCTTAAAGTCCCTCATATGCTTTTATGTGTCAGTTTGTGATAGGTTATTTGT 1860
Db 1801 GTGTGAATTTGCTTAAAGTCCCTCATATGCTTTTATGTGTCAGTTTGTGATAGGTTATTTGT 1860
QY 1861 CATAGAATTTATTTCTTATTTCTTACATTTTATGATTTACTATGATGTATGAGAAATAACACCTA 1920
Db 1861 CATAGAATTTATTTCTTATTTCTTACATTTTATGATTTACTATGATGTATGAGAAATAACACCTA 1920
QY 1921 ATCCCTTATACCTTACCTCAATTTTAACTCCCTTTTAAAGAACTTACATTTACAGATAAAGA 1980
Db 1921 ATCCCTTATACCTTACCTCAATTTTAACTCCCTTTTAAAGAACTTACATTTACAGATAAAGA 1980
QY 1981 TTTTAAAAAATATATTTTGTAGAGACAGGCTTTAGCCACGCGAGGCTGCTCTCT 2040
Db 1981 TTTTAAAAAATATATTTTGTAGAGACAGGCTTTAGCCACGCGAGGCTGCTCTCTCTCTCT 2040
QY 2041 AAGTCTGGCCCAAGCGATCCTCCTGGCTGGCCCTCCTTAAAGTGTCTGGAATTTATAGACAT 2100
Db 2041 AAGTCTGGCCCAAGCGATCCTCCTGGCTGGCCCTCCTTAAAGTGTCTGGAATTTATAGACAT 2100
QY 2101 GAGCCATCAGATCCCAATATACAGAAATAAGATTTTAAAGTGTCTGGAATTTATAGACAT 2160
Db 2101 GAGCCATCAGATCCCAATATACAGAAATAAGATTTTAAAGTGTCTGGAATTTATAGACAT 2160
QY 2161 AAAATTTTCTGAGGCTCAGACAAATGTCAAATGTCTCTCAGTTTACACTGAGATTTTGTAA 2220
Db 2161 AAAATTTTCTGAGGCTCAGACAAATGTCAAATGTCTCTCAGTTTACACTGAGATTTTGTAA 2220

QY	2221	AACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTGTTCAAAGTAAA	2280		
Db	2221	AACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTGTTCAAAGTAAA	2280		
QY	2281	ATGGAAGCAAGAGTAAATCAGCAGTTGAATTCAGAGAAAGACAGAAAAGGAGAAAAG	2340		
Db	2281	ATGGAAGCAAGAGTAAATCAGCAGTTGAATTCAGAGAAAGACAGAAAAGGAGAAAAG	2340		
QY	2341	ATGAAATTCACAGACAGAGAGGAAATATATTATCAATTAAGGAGCAGATCTGTAGA	2400		
Db	2341	ATGAAATTCACAGACAGAGAGGAAATATATTATCAATTAAGGAGCAGATCTGTAGA	2400		
QY	2401	GCTCATTTAGTATGGCAAAATGACTTGTCTCAGGATTAATTTTAAACCCGCTTGTCTGGT	2460		
Db	2401	GCTCATTTAGTATGGCAAAATGACTTGTCTCAGGATTAATTTTAAACCCGCTTGTCTGGT	2460		
QY	2461	TTGCACGGCTGGGATGCAGCTAGGGTTCTGCCTCAGGAGCAGACAGCTGTCAGAGCAGC	2520		
Db	2461	TTGCACGGCTGGGATGCAGCTAGGGTTCTGCCTCAGGAGCAGACAGCTGTCAGAGCAGC	2520		
QY	2521	TGTCAGGCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCTACTCAGGACAGAAAT	2580		
Db	2521	TGTCAGGCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCTACTCAGGACAGAAAT	2580		
QY	2581	GACGAGAACAGGGAGCTGGAACAGAGCCCTTAACAGAGAGAGGAAATGGATCAACA	2640		
Db	2581	GACGAGAACAGGGAGCTGGAACAGAGCCCTTAACAGAGAGAGGAAATGGATCAACA	2640		
QY	2641	AAGTTAACTACAGGTCAGGATCAGGCAATTCATTTCACTCTGACTGCTAATCATGTGACA	2700		
Db	2641	AAGTTAACTACAGGTCAGGATCAGGCAATTCATTTCACTCTGACTGCTAATCATGTGACA	2700		
QY	2701	GAACAGTGTAGGCTTATGTATTTTCATGTAGAGTAGGACCCAAATAATCCACCCAAAGT	2760		
Db	2701	GAACAGTGTAGGCTTATGTATTTTCATGTAGAGTAGGACCCAAATAATCCACCCAAAGT	2760		
QY	2761	CCTTTATCTATGCCACATCCTTCTATCTATCTACTTCCAGGACACTTTTCTTCCCTATGA	2820		
Db	2761	CCTTTATCTATGCCACATCCTTCTATCTATCTACTTCCAGGACACTTTTCTTCCCTATGA	2820		
QY	2821	TAAGGCTCTCTCTCTCCA	2880		
Db	2821	TAAGGCTCTCTCTCTCCA	2880		
QY	2881	CACAAACACACACCCCGCCAAACAGGTGCATGTAAAAGATGTAGATTCCTCTGCCCTTT	2940		
Db	2881	CACAAACACACACCCCGCCAAACAGGTGCATGTAAAAGATGTAGATTCCTCTGCCCTTT	2940		
QY	2941	CTCATCTACAGCCCCAGGAGGGTAAGTTAATATAAGAGGATTTATTGGTAAGAGATGA	3000		
Db	2941	CTCATCTACAGCCCCAGGAGGGTAAGTTAATATAAGAGGATTTATTGGTAAGAGATGA	3000		
QY	3001	TGCTTAATCTGTTTAACTGGGCTCAAAGAGAAATTCCTTCTCTGTACTTATTA	3060		
Db	3001	TGCTTAATCTGTTTAACTGGGCTCAAAGAGAAATTCCTTCTCTGTACTTATTA	3060		
QY	3061	AGCACCTATTATGTCTGAGCTTATATACAAAGGGTTATTATATGCTAATATAGTAAT	3120		
Db	3061	AGCACCTATTATGTCTGAGCTTATATACAAAGGGTTATTATATGCTAATATAGTAAT	3120		
QY	3121	AGTAATGTGTTGGTACTATGTTAAATTTACCATAAAAAATTAATATCCCTTTTAAATAAAG	3180		
Db	3121	AGTAATGTGTTGGTACTATGTTAAATTTACCATAAAAAATTAATATCCCTTTTAAATAAAG	3180		
QY	3181	CTAATTAATATGGAATCTTTTATGATTCATTTTATGTTTATGTTTATGTTTATGTTT	3240		
Db	3181	CTAATTAATATGGAATCTTTTATGATTCATTTTATGTTTATGTTTATGTTTATGTTT	3240		
QY	3241	AAAAGACAACTCTACCCCTGTACCCAGGCTGGAGTGCAGTGGTCAATCATAGCTTCTG	3300		
Db	3241	AAAAGACAACTCTACCCCTGTGTACCCAGGCTGGAGTGCAGTGGTCAATCATAGCTTCTG	3300		
QY	3301	CAGTCTTGAACCTCCTGGGCTCAAGCAATCCTCCTGGCTCCCAAGTGTGGGAT	3360		
Db	3301	CAGTCTTGAACCTCCTGGGCTCAAGCAATCCTCCTGGCTCCCAAGTGTGGGAT	3360		
QY	3361	ACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAATATGATTTTAA	3420		
Db	3361	ACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAATATGATTTTAA	3420		
QY	3421	TTTTTAAATAATATATGCCTTAAATTTTACCTTATGTATGTATCTGTTGTAATAAATCTAG	3480		
Db	3421	TTTTTAAATAATATATGCCTTAAATTTTACCTTATGTATGTATCTGTTGTAATAAATCTAG	3480		
QY	3481	TTTGTGCTCTAAAGTTTAAAGTCTTTCCTTCAATAAGCTTCATGTAGTGGAGGAGACATTT	3540		
Db	3481	TTTGTGCTCTAAAGTTTAAAGTCTTTCCTTCAATAAGCTTCATGTAGTGGAGGAGACATTT	3540		
QY	3541	TAAAGTGAACACAGACAGCAGGTGTGGCTCAGCCCTGTAATCCCAGCACTCTGGGAG	3600		
Db	3541	TAAAGTGAACACAGACAGCAGGTGTGGCTCAGCCCTGTAATCCCAGCACTCTGGGAG	3600		
QY	3601	GCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCCTGAGCAACATGGCAA	3660		
Db	3601	GCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCCTGAGCAACATGGCAA	3660		
QY	3661	AACCTCTTCTTATTAACAAAAATTAGCCGGGATGGTGGCATGTGCTGTGCTGCCAGCT	3720		
Db	3661	AACCTCTTCTTATTAACAAAAATTAGCCGGGATGGTGGCATGTGCTGTGCTGCCAGCT	3720		
QY	3721	ACTAGGGGCTGAGGACGAGAGATCTTTGGAGCCAGGAGTCAAGGCTGCACTGAGCAG	3780		
Db	3721	ACTAGGGGCTGAGGACGAGAGATCTTTGGAGCCAGGAGTCAAGGCTGCACTGAGCAG	3780		
QY	3781	TGCTTGCCCACTGCACCTCCAGCCTGGGTGACAGCAGCAGACTTTGCCCTCAAAAAATAA	3840		
Db	3781	TGCTTGCCCACTGCACCTCCAGCCTGGGTGACAGCAGCAGACTTTGCCCTCAAAAAATAA	3840		
QY	3841	GAAGAAAAATTTAAAAATTAATGGAACAACACTACAAAGAGCTGTTGCTCTAGATGAGCTAC	3900		
Db	3841	GAAGAAAAATTTAAAAATTAATGGAACAACACTACAAAGAGCTGTTGCTCTAGATGAGCTAC	3900		
QY	3901	TTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTAC	3960		
Db	3901	TTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTAC	3960		
QY	3961	ATTATTAATAATCAATTTCTCAATGTATATCCACAAAGAGCTGGTACCTGATGTTCAT	4020		
Db	3961	ATTATTAATAATCAATTTCTCAATGTATATCCACAAAGAGCTGGTACCTGATGTTCAT	4020		
QY	4021	AGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080		
Db	4021	AGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080		
QY	4081	CAATTAACAAAAATGCTATATATCCATGCAATGGAATACCCACCTGCACTACAAAGGAAG	4140		
Db	4081	CAATTAACAAAAATGCTATATATCCATGCAATGGAATACCCACCTGCACTACAAAGGAAG	4140		
QY	4141	AAGTACTTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCAGACATCAAGG	4200		
Db	4141	AAGTACTTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCAGACATCAAGG	4200		
QY	4201	AGGAGATTAATGTATGCCATACGAAATTTCTAGAAAAATGAAGTAACCTATATAGTTACAGAAA	4260		
Db	4201	AGGAGATTAATGTATGCCATACGAAATTTCTAGAAAAATGAAGTAACCTATATAGTTACAGAAA	4260		
QY	4261	GCAATCAGGGCAGGCATAGAGGCTCACACCTGTATATCCAGCACTTTGAGAGGCCACGT	4320		
Db	4261	GCAATCAGGGCAGGCATAGAGGCTCACACCTGTATATCCAGCACTTTGAGAGGCCACGT	4320		
QY	4321	GGGAAGATTGCTAGAACTCAGGAGTTCAAGACAGCCCTGGGCAACACAGTGAACCTCCAT	4380		
Db	4321	GGGAAGATTGCTAGAACTCAGGAGTTCAAGACAGCCCTGGGCAACACAGTGAACCTCCAT	4380		
QY	4381	TCTCCAAAAAATGGAAAAAAGAAAGCAATCAAGTGGTGTCTCTGCTGGGAGGGGAAG	4440		
Db	4381	TCTCCAAAAAATGGAAAAAAGAAAGCAATCAAGTGGTGTCTCTGCTGGGAGGGGAAG	4440		

```

Db 4381 TCTCACAAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTGTCTCTGTGGGAGGGAAG 4440
QY 4441 GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGTGGTGAATTCAGGTCTCTATCCT 4500
Db 4441 GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGTGGTGAATTCAGGTCTCTATCCT 4500
QY 4501 GACTGTGTACAGTTTGGGGTGTTCATCCAAAAATATTCGTPAGAATATGTCATCTTA 4560
Db 4501 GACTGTGTACAGTTTGGGGTGTTCATCCAAAAATATTCGTPAGAATATGTCATCTTA 4560
QY 4561 AATGGGTGGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAAATAATGTGTAAG 4620
Db 4561 AATGGGTGGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAAATAATGTGTAAG 4620
QY 4621 AAAAGTTTCAATCTCTGCGCAGCAACGTTATTCAAATTCCTGAGCCCTTTACTCGCA 4680
Db 4621 AAAAGTTTCAATCTCTGCGCAGCAACGTTATTCAAATTCCTGAGCCCTTTACTCGCA 4680
QY 4681 AATTCCTGCACTTCTGCGCCGTTACCATAGGTAGGTACAGCACTAGCTCCACAAATGGATA 4740
Db 4681 AATTCCTGCACTTCTGCGCCGTTACCATAGGTAGGTAGGTAGGTAGGTAGGTAGGTAG 4740
QY 4741 AATGCAATTTCTGGAAGAGACTAGGCAAAATCCAGCATCACTTGTGCTTCAATATCA 4800
Db 4741 AATGCAATTTCTGGAAGAGACTAGGCAAAATCCAGCATCACTTGTGCTTCAATATCA 4800
QY 4801 CCAGCGTGTACAGCTTGTGTGCTGTGCGAGTGCATGCAATGGGAGCTCTTGAATTCCTTAA 4860
Db 4801 CCAGCGTGTACAGCTTGTGTGCTGTGCGAGTGCATGCAATGGGAGCTCTTGAATTCCTTAA 4860
QY 4861 GGAACCTTGGGTTTACCAGAGTATTTCCACAAATGCTATTCAATAGTCTTATGATAG 4920
Db 4861 GGAACCTTGGGTTTACCAGAGTATTTCCACAAATGCTATTCAATAGTCTTATGATAG 4920
QY 4921 CAAGACACTGTGCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGA 4980
Db 4921 CAAGACACTGTGCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGA 4980
QY 4981 ACAACATAGCAAGATATTTAGATCATTTTGAATGCTTAAAGAGCAGAGTACAAAT 5040
Db 4981 ACAACATAGCAAGATATTTAGATCATTTTGAATGCTTAAAGAGCAGAGTACAAAT 5040
QY 5041 CACACATGCAATCAGTATATCCAAATCATGTAATATGTCCTGTAGAAAGCAGAGG 5100
Db 5041 CACACATGCAATCAGTATATCCAAATCATGTAATATGTCCTGTAGAAAGCAGAGG 5100
QY 5101 AATAACACAGAAATCTTAACAGTCAATGCTATTAGACACTTAAGTCTAATATTATTT 5160
Db 5101 AATAACACAGAAATCTTAACAGTCAATGCTATTAGACACTTAAGTCTAATATTATTT 5160
QY 5161 AGACACTATGATATTGAGATTTAAAAATCTTTAATATTTTAAATTTAGAGCTCTCT 5220
Db 5161 AGACACTATGATATTGAGATTTAAAAATCTTTAATATTTTAAATTTAGAGCTCTCT 5220
QY 5221 ATTTTTCATAGTATTCAAGTTTGAATGATCAAGTATTAAGTATTAAGTATTAAGTAT 5280
Db 5221 ATTTTTCATAGTATTCAAGTTTGAATGATCAAGTATTAAGTATTAAGTATTAAGTAT 5280
QY 5281 TTTTTTTTTTTTGGAGTGGAGTTTGGTCTGTGTCCTGCTGCTGCTGCTGCTGCTGCTG 5340
Db 5281 TTTTTTTTTTTTGGAGTGGAGTTTGGTCTGTGTCCTGCTGCTGCTGCTGCTGCTGCTG 5340
QY 5341 AYCATACTCACTGCAACCTCCACTCTGGGTTTCAAGCAAAAGTGTGCGCTCAGCCTCC 5400
Db 5341 AYCATACTCACTGCAACCTCCACTCTGGGTTTCAAGCAAAAGTGTGCGCTCAGCCTCC 5400
QY 5401 CGGTAGATGGATTTACAGGCGCCACCAACACACTCGGCTAATGTTGATTTTAGTA 5460
Db 5401 CGGTAGATGGATTTACAGGCGCCACCAACACACTCGGCTAATGTTGATTTTAGTA 5460
QY 5461 GAGATGGGTTTACCAATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAGAG 5514
Db 5461 GAGATGGGTTTACCAATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAGAG 5514

```

```

RESULT 2
US-09-966-880A-35
: Sequence 35, Application US/09966880A
: Patent No. US20020164743A1
: GENERAL INFORMATION:
: APPLICANT: Honjo, Tasuku
: TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
: FILE REFERENCE: 06501-088001
: CURRENT APPLICATION NUMBER: US/09/966,880A
: CURRENT FILING DATE: 2001-09-28
: PRIOR APPLICATION NUMBER: PCT/JP00/01918
: PRIOR FILING DATE: 2000-03-28
: PRIOR APPLICATION NUMBER: JP 11-371382
: PRIOR FILING DATE: 1999-12-27
: PRIOR APPLICATION NUMBER: JP 11-178999
: PRIOR FILING DATE: 1999-06-24
: PRIOR APPLICATION NUMBER: JP 11-87192
: PRIOR FILING DATE: 1999-03-29
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 35
: LENGTH: 11204
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-966-880A-35

```

```

Query Match      89.3%; Score 4922.8; DB 9; Length 11204;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4921; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 591 AGGTTTCAGAGAGCTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTTCAACAC 650
Db 1 AGGTTTCAGAGAGCTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTTCAACAC 60
QY 651 AATAACCCCAAGAGCTATTTAAATGCTCTTTAAAGGTATTTACATAAATATTAATATTCTC 710
Db 61 AATAACCCCAAGAGCTATTTAAATGCTCTTTAAAGGTATTTACATAAATATTAATATTCTC 120
QY 711 ATTGTGCTTTTATTTTGTGTATCATGATTATATAATGAAGTGTCTACTGTACTGCTCC 770
Db 121 ATTGTGCTTTTATTTTGTGTATCATGATTATATAATGAAGTGTCTACTGTACTGCTCC 180
QY 771 TGATCTTTGCTAGCTATGGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCAAGGAACC 830
Db 181 TGATCTTTGCTAGCTATGGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCAAGGAACC 240
QY 831 TAAACATTAAGCAGAGCTGCCCTCAATGTTTAACTGTGTGACTCTGCTCATGACAGC 890
Db 241 TAAACATTAAGCAGAGCTGCCCTCAATGTTTAACTGTGTGACTCTGCTCATGACAGC 300
QY 891 CCCACCCACCATCTTCACTGGATCCAAATCAGGAGCAAGCCGTTGGGTACCTGGTGG 950
Db 301 CCCACCCACCATCTTCACTGGATCCAAATCAGGAGCAAGCCGTTGGGTACCTGGTGG 360
QY 951 GGTGTGTGTCTCAGGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGCC 1010
Db 361 GGTGTGTGTCTCAGGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGCC 420
QY 1011 AATGCACTGTACAGCTGAGCAGAGAACCATCATTAATTTGAATGAGATTTTCTGGCT 1070
Db 421 AATGCACTGTACAGCTGAGCAGAGAACCATCATTAATTTGAATGAGATTTTCTGGCT 480
QY 1071 GAGACTTTCAGGAGGAGCAAGAGACACTCTGACACCACTATGAGCAGTAAAGAGGAG 1130
Db 481 GAGACTTTCAGGAGGAGCAAGAGACACTCTGACACCACTATGAGCAGTAAAGAGGAG 540
QY 1131 TCTTCTCGTGGTGTGATGTCAGCTGGGCTTCTCTCAGAGCAAAATCTGAGTAAATGAGACTGG 1190
Db 541 TCTTCTCGTGGTGTGATGTCAGCTGGGCTTCTCTCAGAGCAAAATCTGAGTAAATGAGACTGG 600

```


Qy	1191	TAGCTATCCCTTTCTCTCATGTAACTGTCTGACGTGATAAGATCAGCTTGATCAATATGCA	1250
Db	601	TAGCTATCCCTTTCTCTCATGTAACTGTCTGACGTGATAAGATCAGCTTGATCAATATGCA	660
Qy	1251	TATATATTTTGTGATCTGTCTCCTTTCTCTCTATTCAGATCTTATAGCTGTACGCCAA	1310
Db	661	TATATATTTTGTGATCTGTCTCCTTTCTCTCTATTCAGATCTTATAGCTGTACGCCAA	720
Qy	1311	TTCTTTCTGTTCAGACTCTCTTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTG	1370
Db	721	TTCTTTCTGTTCAGACTCTCTTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTG	780
Qy	1371	CGTAAATGACTGATTCGCTCGCTGAGATTTGTACCATGGTGAAGTAAATTTATGGTAAT	1430
Db	781	CGTAAATGACTGATTCGCTCGCTGAGATTTGTACCATGGTGAAGTAAATTTATGGTAAT	840
Qy	1431	AATATTAACTAGCAAACTTTTAGAGACTCAAAATCATGAAGAAGTAATAGCAGTACTGTA	1490
Db	841	AATATTAACTAGCAAACTTTTAGAGACTCAAAATCATGAAGAAGTAATAGCAGTACTGTA	900
Qy	1491	CTAAAAACGGTAGTGCCTAAATTTTCGTAAATTAATTTTGTAAATATTTCAACAGTAAACCAACT	1550
Db	901	CTAAAAACGGTAGTGCCTAAATTTTCGTAAATTAATTTTGTAAATATTTCAACAGTAAACCAACT	960
Qy	1551	TGAAGACACACTTTCCCTAGGAGCGCTTACTGAAATAATTTAGCTATAGTAAGAAAAATTT	1610
Db	961	TGAAGACACACTTTCCCTAGGAGCGCTTACTGAAATAATTTAGCTATAGTAAGAAAAATTT	1020
Qy	1611	GTAAATTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATGTGTC	1670
Db	1021	GTAAATTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATGTGTC	1080
Qy	1671	CATTATAGGAGACAAATTCATTCACGCAAGTATTTAAATGTTAAAGGCCCAATTTGTAG	1730
Db	1081	CATTATAGGAGACAAATTCATTCACGCAAGTATTTAAATGTTAAAGGCCCAATTTGTAG	1140
Qy	1731	GCAGTTAATGGCACTTTTACTATTAATTAATTTTCCATTTGTTCCACTTTGTTGACAGCTTAACT	1790
Db	1141	GCAGTTAATGGCACTTTTACTATTAATTAATTTTCCATTTGTTCCACTTTGTTGACAGCTTAACT	1200
Qy	1791	TACCTCTTAGTGGAATTTGGTTAAGGTCCTCATATGCTTTTATGTGCAAGTTTGTAT	1850
Db	1201	TACCTCTTAGTGGAATTTGGTTAAGGTCCTCATATGCTTTTATGTGCAAGTTTGTAT	1260
Qy	1851	AGGTTATGTGCATAGAACTTATCTATTCCTACATTTATGATTTACTATGGATGTATGAGA	1910
Db	1261	AGGTTATGTGCATAGAACTTATCTATTCCTACATTTATGATTTACTATGGATGTATGAGA	1320
Qy	1911	ATAACACCTAATCCCTTATACCTTACCTCAATTTTAACTCCCTTTATAAAGAACTTACATAC	1970
Db	1321	ATAACACCTAATCCCTTATACCTTACCTCAATTTTAACTCCCTTTATAAAGAACTTACATAC	1380
Qy	1971	AGATAAGATTTTTTAAAAATATATTTTGTGTAGACACAGGTCCTTAGCCAGCCGAG	2030
Db	1381	AGATAAGATTTTTTAAAAATATATTTTGTGTAGACACAGGTCCTTAGCCAGCCGAG	1440
Qy	2031	GCTGGTCTCTAAGTCTCTGGCCCAAGCGATCCTCGCTGGCCCTCTCTAAAGTCTGGAA	2090
Db	1441	GCTGGTCTCTAAGTCTCTGGCCCAAGCGATCCTCGCTGGCCCTCTCTAAAGTCTGGAA	1500
Qy	2091	TTATAGACATGAGCCATCACATCCAAATATACAGAAATAAGATTTTTTATGGAGATTAA	2150
Db	1501	TTATAGACATGAGCCATCACATCCAAATATACAGAAATAAGATTTTTTATGGAGATTAA	1560
Qy	2151	TGTTCTTCAGAAAAATTTTCTGAGGTCAGACAAATGTCAAATGTCCTCAGTTTACACTG	2210
Db	1561	TGTTCTTCAGAAAAATTTTCTGAGGTCAGACAAATGTCAAATGTCCTCAGTTTACACTG	1620
Qy	2211	AGATTTTTGAAACAAAGTCTGAGCTATAGGTCCTTGTGAAGGCTCCATTTGGAATACTTTGT	2270
Db	1621	AGATTTTTGAAACAAAGTCTGAGCTATAGGTCCTTGTGAAGGCTCCATTTGGAATACTTTGT	1680
Qy	2271	TCAAAGTAAATGGAAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAA	2330

Db	1681	TCAAAGTAAATGGAAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAA	1740
Qy	2331	AGGAGAAAGATGAAATTTCAACAGACAGAGAGGGAATATATTTATCATTAAGGAGACAG	2390
Db	1741	AGGAGAAAGATGAAATTTCAACAGACAGAGAGGGAATATATTTATCATTAAGGAGACAG	1800
Qy	2391	TATCTGTAGAGCTCATTTAGTGTGATGCAAAATGACTTGTGTCAGGATTTATTTTAAACCCGT	2450
Db	1801	TATCTGTAGAGCTCATTTAGTGTGATGCAAAATGACTTGTGTCAGGATTTATTTTAAACCCGT	1860
Qy	2451	TGTTTCTGTGTTGCACGGCTGGGATGCAGCTAGGAGTTGTCCTCAGGAGACAGAGCTGT	2510
Db	1861	TGTTTCTGTGTTGCACGGCTGGGATGCAGCTAGGAGTTGTCCTCAGGAGACAGAGCTGT	1920
Qy	2511	CCAGAGCAGCTGTGAGCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCTACTCA	2570
Db	1921	CCAGAGCAGCTGTGAGCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCTACTCA	1980
Qy	2571	GGACAGAAATCACGAGAACAGGAGCTGGAACAGGCCCTTAACACAGAGAGGGAAGTAA	2630
Db	1981	GGACAGAAATCACGAGAACAGGAGCTGGAACAGGCCCTTAACACAGAGAGGGAAGTAA	2040
Qy	2631	TGGATCAACAAGTTAATCTAGCAGGTCAAGGATCAGCAATTCATTTCTACTCTGACTGTA	2690
Db	2041	TGGATCAACAAGTTAATCTAGCAGGTCAAGGATCAGCAATTCATTTCTACTCTGACTGTA	2100
Qy	2691	ACATGTGACAGAAACAGTGTAGGCTTATTTGATTTTTCATGTAGAGTAGGACCCCAAAATC	2750
Db	2101	ACATGTGACAGAAACAGTGTAGGCTTATTTGATTTTTCATGTAGAGTAGGACCCCAAAATC	2160
Qy	2751	CACCCAAAGTCTTTATCTATGCCACATCTTCTTATCTATCTACTTCCAGGACACTTTTTC	2810
Db	2161	CACCCAAAGTCTTTATCTATGCCACATCTTCTTATCTATCTACTTCCAGGACACTTTTTC	2220
Qy	2811	TTCTCTTATGATAAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2870
Db	2221	TTCTCTTATGATAAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2280
Qy	2871	CACACACACACAAACA	2930
Db	2281	CACACACACACAAACA	2340
Qy	2931	CTCTGCCCTTCTCATCTACACAGCCAGGAGGTAAGTTAATATTAAGAGGATTTATTTGG	2990
Db	2341	CTCTGCCCTTCTCATCTACACAGCCAGGAGGTAAGTTAATATTAAGAGGATTTATTTGG	2400
Qy	2991	TAAGAGATGATGCTTAAATCTGTTTAACTGCTGAGGCTCAAGAGAGAAATTTCTTTCTCT	3050
Db	2401	TAAGAGATGATGCTTAAATCTGTTTAACTGCTGAGGCTCAAGAGAGAAATTTCTTTCTCT	2460
Qy	3051	GTACTTATTAAGCACCTAATATGCTTGCAGCTTATATACAAAGGGTTATTTATATGCTA	3110
Db	2461	GTACTTATTAAGCACCTAATATGCTTGCAGCTTATATACAAAGGGTTATTTATATGCTA	2520
Qy	3111	ATATAGTAATAGTAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	3170
Db	2521	ATATAGTAATAGTAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	2580
Qy	3171	TAAATAAAGCTAATATTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	3230
Db	2581	TAAATAAAGCTAATATTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	2640
Qy	3231	TTGATTTTTTAAAGACAACTCACCCTGTTTACCAGGCTGAGTGCAAGGTAAGTCA	3290
Db	2641	TTGATTTTTTAAAGACAACTCACCCTGTTTACCAGGCTGAGTGCAAGGTAAGTCA	2700
Qy	3291	TAGCTTTCTGAGTCTTTGAACCTCTGGGCTCAAGCAATCTCCTCCTTGGCTCCCAAA	3350
Db	2701	TAGCTTTCTGAGTCTTTGAACCTCTGGGCTCAAGCAATCTCCTCCTTGGCTCCCAAA	2760
Qy	3351	GTGTTGGGATACAGTCATGAGCCACTGTCATCTGGCCTAGGATCCATTTAGATTAAATAT	3410

Db	2761	GTGTTGGATACAGTCATGAGCCACTGCATCTGCGCCTAGGATCCATTTAGATTAAATAT	2820
Qy	3411	GCATTTTAAATTTAAATTAATATGGCTAAATTTTACCTTTATGTAAATGTGTATACCTGGTA	3470
Db	2821	GCATTTTAAATTTTAAATTAATATGGCTAAATTTTACCTTTATGTAAATGTGTATACCTGGTA	2880
Qy	3471	ATAAATCTAGTTTGGCTGAAAGTTTAAAGTTTAAAGTGTCTTCCAAATAGCTTCAATGACGTGAG	3530
Db	2881	ATAAATCTAGTTTGGCTGAAAGTTTAAAGTTTAAAGTGTCTTCCAAATAGCTTCAATGACGTGAG	2940
Qy	3531	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGCTCACGCCCTGTATCCACGC	3590
Db	2941	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGCTCACGCCCTGTATCCACGC	3000
Qy	3591	ACTCTGGGAGGCTGAGTGGGTGGATCGCTGTAGCCCTGAGGTTCGAAGACCCAGCCTGAGC	3650
Db	3001	ACTCTGGGAGGCTGAGTGGGTGGATCGCTGTAGCCCTGAGGTTCGAAGACCCAGCCTGAGC	3060
Qy	3651	AACATGGCAAAACCCCTGTTCTATATACAAAAATTTAGCCGGGCATGGTGGCATGTGCTGT	3710
Db	3061	AACATGGCAAAACCCCTGTTCTATATACAAAAATTTAGCCGGGCATGGTGGCATGTGCTGT	3120
Qy	3711	GGTCCAGCTACTAGGGGGCTGAGGCAGGAGAACTTTTGGAGCCCAAGGAGGTCAAGGCTG	3770
Db	3121	GGTCCAGCTACTAGGGGGCTGAGGCAGGAGAACTTTTGGAGCCCAAGGAGGTCAAGGCTG	3180
Qy	3771	CAGTGAGCAGTGTCTGGCCACTGCACTCCAGCTGGGTGACAGGACCAGACCTTGGCCTC	3830
Db	3181	CAGTGAGCAGTGTCTGGCCACTGCACTCCAGCTGGGTGACAGGACCAGACCTTGGCCTC	3240
Qy	3831	AAAAAATTAAGAGAAATTTAAATAAATGGAACAACTACAAAGAGCTTGTTCCTA	3890
Db	3241	AAAAAATTAAGAGAAATTTAAATAAATGGAACAACTACAAAGAGCTTGTTCCTA	3300
Qy	3891	GATGAGCTACTTAGTTAGCTGATATTTTGGTATTATTAACCTTTAAAGTCAGGCTGTGCA	3950
Db	3301	GATGAGCTACTTAGTTAGCTGATATTTTGGTATTATTAACCTTTAAAGTCAGGCTGTGCA	3360
Qy	3951	CTGCACATACATTTAAATATCAATCTCAATGTATATCCACAAGAGCTGTGACGT	4010
Db	3361	CTGCACATACATTTAAATATCAATCTCAATGTATATCCACAAGAGAGCTGTGACGT	3420
Qy	4011	GAATGTTCATAGTACTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	4070
Db	3421	GAATGTTCATAGTACTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	3480
Qy	4071	ACAAGTGAACAAATTAACAAAATGTGCTATATCCATGCAATGGAATACCCACCTGCACT	4130
Db	3481	ACAAGTGAACAAATTAACAAAATGTGCTATATCCATGCAATGGAATACCCACCTGCACT	3540
Qy	4131	ACAAAGGAAGAGCTTACTTGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCA	4190
Db	3541	ACAAAGGAAGAGCTTACTTGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCA	3600
Qy	4191	GACATGAAGAGGAGATATGTATGCCATACGAAATCTAGAAAATGAAAGTAACTTATA	4250
Db	3601	GACATGAAGAGGAGATATGTATGCCATACGAAATCTAGAAAATGAAAGTAACTTATA	3660
Qy	4251	GTTACAGAAACCAATCAGGCGAGGCATAGAGGCTCACACTGTAAATCCCAAGCACTTTGA	4310
Db	3661	GTTACAGAAACCAATCAGGCGAGGCATAGAGGCTCACACTGTAAATCCCAAGCACTTTGA	3720
Qy	4311	GAGGCCACGTGGGAAGATTGCTAGAACTCAGAGGTTCAAGACCCAGCCTGGGCAACACAGT	4370
Db	3721	GAGGCCACGTGGGAAGATTGCTAGAACTCAGAGGTTCAAGACCCAGCCTGGGCAACACAGT	3780
Qy	4371	GAATCTCATCTCCACAAAATGGGAAAAAGAAAAGCAAAATCAGTGTGCTGCTGG	4430
Db	3781	GAATCTCATCTCTCCACAAAATGGGAAAAAGAAAAGCAAAATCAGTGTGCTGCTGG	3840
Qy	4431	GGAGGGAGGACTGCAAGAGGGAACAGCTCTGTGGGTGAGGTGATTCAGGT	4490
Db	3841	GGAGGGAGGACTGCAAGAGGGAAGAGCTCTGTGGGTGAGGTGATTCAGGT	3900
Qy	4491	TCTGTATCTCTGACTGTGGTAGCAGTTTGGGTGCTTTACATCCAAAAATATTTCGTAGAAAT	4550
Db	3901	TCTGTATCTCTGACTGTGGTAGCAGTTTGGGTGCTTTACATCCAAAAATATTTCGTAGAAAT	3960
Qy	4551	ATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTTACTCTCAATGTAAAGAAAAAT	4610
Db	3961	ATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTTACTCTCAATGTAAAGAAAAAT	4020
Qy	4611	AATGTGTAAAGAAAAAGTTTCAATTCCTTCCAGCAAAACGTTATTCAAAATTCCTGAGCCCT	4670
Db	4021	AATGTGTAAAGAAAAAGTTTCAATTCCTTCCAGCAAAACGTTATTCAAAATTCCTGAGCCCT	4080
Qy	4671	TTACTTTCGCAAAATCTCTGCACCTTCTGCCCCGTACCATTTAGGTGACAGACACTAGCTCCAC	4730
Db	4081	TTACTTTCGCAAAATCTCTGCACCTTCTGCCCCGTACCATTTAGGTGACAGACACTAGCTCCAC	4140
Qy	4731	AAATTTGGATTAATGCAATTTCTGGAAAGACTAGGACAAAAATCCAGGCATCACTTTGTGCT	4790
Db	4141	AAATTTGGATTAATGCAATTTCTGSAAGAACTAGGACAAAAATCCAGGCATCACTTTGTGCT	4200
Qy	4791	TTTCATATCAACACACGCTGTACAGCTTGTGCTGTCTGACAGTCAATTTGGAAGACTCTTG	4850
Db	4201	TTTCATATCAACACACGCTGTACAGCTTGTGCTGTCTGACAGTCAATTTGGAAGACTCTTG	4260
Qy	4851	ATTTCTTTAAGGAAACTTTGGGTTACAGAGTATTTCCCAAAATGCTATTCAANTTAGTGC	4910
Db	4261	ATTTCTTTAAGGAAACTTTGGGTTACAGAGTATTTCCCAAAATGCTATTCAANTTAGTGC	4320
Qy	4911	TTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCAGTCA	4970
Db	4321	TTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCAGTCA	4380
Qy	4971	TTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGTACTAGTTTAAAAAGCAGCAG	5030
Db	4381	TTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGTACTAGTTTAAAAAGCAGCAG	4440
Qy	5031	AGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGCGCTGACAA	5090
Db	4441	AGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGCGCTGACAA	4500
Qy	5091	AGACTAGAGGAATTAACACAAGAACTTTTAACAGTCAATTTAGACACTAAAGTCTAAAT	5150
Db	4501	AGACTAGAGGAATTAACACAAGAACTTTTAACAGTCAATTTAGACACTAAAGTCTAAAT	4560
Qy	5151	TATTATTATTAGACACTATGATATTTGAGATTTTAAAAAATCTTTTAAATTTTAAAAATTTA	5210
Db	4561	TATTATTATTAGACACTATGATATTTGAGATTTTAAAAAATCTTTTAAATTTTAAAAATTTA	4620
Qy	5211	GAGCTCTCTATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGTATTACTCTTTCTT	5270
Db	4621	GAGCTCTCTATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGTATTACTCTTTCTT	4680
Qy	5271	TTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATGAGTTTGGCTGTGTGGCCCATGCTGGAGTG	5330
Db	4681	TTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATGAGTTTGGCTGTGTGGCCCATGCTGGAGTG	4740
Qy	5331	GAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCCTGGTTCACAGCAAGCTGTCGC	5390
Db	4741	GAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGTTCACAGCAAGCTGTCGC	4800
Qy	5391	CTCAGCCTCCGGGTAGATGGGATTACAGGGCCCCACACACACTCGGCTAATGTGTTGT	5450
Db	4801	CTCAGCCTCCGGGTAGATGGGATTACAGGGCCCCACACACACTCGGCTAATGTGTTGT	4860
Qy	5451	ATTTTATTAGTAGAGTGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCTGACCTC	5510
Db	4861	ATTTTATTAGTAGAGTGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCTGACCTC	4920
Qy	5511	AGAG 5514	
Db	4921	AGAG 4924	

Db	18374	AGATATTACAGTACCGGTATATATGTGTGTGCATATGTGTGCAAGACATATATACCAGTTA	18315
Qy	5019	AAAAAGCAGCAGAGTACAAATACACATCAATCAATCAATCAATCAATCAATCAATCAAT	5078
Db	18314	TCATTACATAGTGGGTATAGGCTAATATAAATGTTTATCTTTGGGCTCTCTATCTTT	18255
Qy	5079	GTCCCTGTAGAAAGACTAGAGGAATAAACACAAGAATCTTAACAGCTATGTCAATAGAC	5138
Db	18254	-TCCAAGTTTCTTTAAAGCCCTGTATCTTTTACATTTGGGGTCAAGGCTGTGTTT	18196
Qy	5139	ACTAAGTCTAATATTATTATTATAGACACTATGATATTGAGATTTAAAAAATCTTTAATA	5198
Db	18195	ATTTTATTATTATTATTATTATTACAGATGCCCACTATAATAATGTTATATATGCGCTGAA	18136
Qy	5199	TTTTAAAAATTTAGAGCTCTCTATTTTTCCATAGTATTCAAAGTTTGACA-----	5247
Db	18135	TTTGCTAAGAGAGTAAACCTGTTTTTTCTTTTGCACTCTCTTTTGTATTTTAATTT	18076
Qy	5248	-----ATGATCAAGTATTACTCTTTCTTTT	5274
Db	18075	TGTGGGTATATAATAGTTGCTTTTATTATTATTATTATTATTATTATTATTATTATTAT	18016
Qy	5275	TTTTTTTTTTTTTTTTTTTGGAGTGGAGTTTTGGTCTCTGTCGCCATGCTGGAGTGAAT	5334
Db	18015	TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT	17956
Qy	5335	GGCATGAYCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAGAGCTGTCGCTCA	5394
Db	17955	GGCGAATCTCAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAATTTCTCCTGCCA	17896
Qy	5395	GCTCCCGGTAGATGGGATACAGCGCCGACCCACCACTCCAGCTCCGCTGTAATTT	5454
Db	17895	GCTCCGCTGAGTGGGACTACAGCGCCGACCCACCACTCCAGCTCCGCTGTAATTT	17836
Qy	5455	TTAGTAGAGATGGGTTTCAACATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA	5511
Db	17835	TTAGTAGAGATGGGTTTCAACATGTTGGCCAGGCTGGTCTGCAACTCCTGACCTCA	17779

RESULT 4

US-09-764-872-517/c
; Sequence 517, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 517
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-872-517

Query Match	4.1%;	Score 226.8;	DB 9;	Length 32204;
Best Local Similarity	48.8%;	Pred. No. 5.8e-37;		
Matches	976;	Conservative 0;	Mismatches 967;	Indels 59; Gaps 11;
Qy	3542	AAAGTGAACACAGACAGCAGGCTGGTGGCTCAGCGCTGTAATCCAGCACTCTGGGAGG	3601	
Db	18793	AAATTAATAAATAAATGCGAGTGGTGGCTCAGACCTGTAATCCAGCACTCTGGGAGG	18734	
Qy	3602	CTGAGTGGGTGGATCGCTTGAGCCCTGGAGTTCAGACCACTGAGCAACATGGCAAA	3661	
Db	18733	CCAAAGCGGTGGATCACTCAGGTGAGGAGTTCAGACCACTGAGCAACATGGTGA	18674	
Qy	3662	ACCTGCTTCTC-----ATAACAAAAATAGCCGGCATGGTGCATGTGCTGTGCTC	3715	
Db	18673	ACCTGCTTCTCTACTAAAAAATACAAAAATAGCTGGGCGGTGGTGTGGGCATCTCTAATCC	18614	

Qy	3716	CAGCTACT-AGGGGGCTGAGGCAGGAGAATCTTTTGGAGCCAGGAGGTCAAGGCTCGACT	3774
Db	18613	CAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCTTGGAGCCAGGAGGAGGTTGCGAGT	18554
Qy	3775	GAGCAGTGTGTGGCCACTGCACCTCCAGCTGGGTGACAGGACACACACTTGCTCTCAAA	3834
Db	18553	AAAGCTGAGATGATGCCACTGCACCTCCAGCTGGAAGACAGAGTGAAGTCTGTCTCAAA	18494
Qy	3835	AAATAAGAAGAAAAATTAATAATAAATGAAACAACTACAAAGAGCTGTGTCTAGATG	3894
Db	18493	ACAAAAACAAACAAACAAACAAATACAGATGGCCAGGCGGATGCTGGCTCACACC	18434
Qy	3895	AGCTACTTATAGTGGCTGA-----TATTTTGGTATTTAATCTTTTAAAGTCAAGGCTGTG	3949
Db	18433	TGTAATACAGCACTATGGAAGGCTGAGGAGGAGTGTCTTAAGCCCTAGAGTTCAAGA	18374
Qy	3950	ACCTGCACCTACATTTAATAATATCAATTTCTCAATGTATATCCACACAAGACTGGTAGG	4009
Db	18373	CCAGAGTGGGCAATGTAGTGAGACCCCCCATCTCTACAGGAACACATAAAATACANATA	18314
Qy	4010	TGAATGTTTCATAGTACCTTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATPCCAT	4069
Db	18313	AAATGGCAGTTTCTCAGTTTGTGCTGGTTCATGTGGTCACTGTCTCTTCAGAAATC	18254
Qy	4070	CAACAAGTGAACAAATAAACAATAATGCTATATCCATGCAATGGAATACCACTGTCAG	4129
Db	18253	ACTGATGGCCAGGTTCACTCTTCAAGACTGGGAAGCAGCAGCAGCAAGGATGGGATTTG	18194
Qy	4130	TACAAAGGAAGAGTACTTTGGGATGAATCCCAAGTCAATGACCTGTAATCAAAAGATC	4189
Db	18193	AGGCCAGAACTCCAGGTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	18134
Qy	4190	AGACATGAAGGAGAGATATGATGCCATAGCAATTTCTAGAAAATCAAAATGAATTAAT	4249
Db	18133	ACGGTGGCTCACACCTGTAATTTCCGGTACTTTGGGAGAACATGGCAAAACCCCTGTCT	18074
Qy	4250	AGTTACAGAAAGCAATCAGGCGAGCATAGAGGCTCACACTGTAATCCAGCACTTTG	4309
Db	18073	CTAAAAATACAAAAAATCAGGCTGGCGGTAGCTCATGCTGTATATCCAGCACTTTG	18014
Qy	4310	AGAGGCCAGT-GGGAAGATGCTAGAACTCAGGAGTTCAAGACAGCCTGGGCAACACA	4368
Db	18013	GGAGGCCAAGGAGGTGGATCACCCTGAGATCAGGAGTCTGAGCCAGCCTGGCCACGCT	17954
Qy	4369	GTAAACTCCATTTCCACAAAAATGGGAAAAAGAAAGCAATCAGTGTGTGCTGT	4428
Db	17953	GCGAAACCCCATCTCTACTAAAAATACAAAAATAGCCGGCATGGTTCTGTACACCTTT	17894
Qy	4429	-----GGGAGGGGAGGAGTGCAAAGAGGGAAGAGTCTGTGGTGGGTGAGGTTGGTAT	4484
Db	17893	AATCCAGCTACTAAGGAGGCTGAGGCAAGAGAATTTGTTGAACCTGGCAAGCGGTGTT	17834
Qy	4485	TCAGTGTCTGTATCTCTGCTGTGTAGCAGTTTGGGTGTTTACATCCAAAAATATTCGT	4544
Db	17833	CGAGTAAGTCAAGATTTGTGCCAATGCAATCCAGCCTGGGTGACAGCAAGCACTCGCT	17774
Qy	4545	AGAATTATGCACTTTAAATGGGTGGAGTTTACTGTATGTAATTTATACCTCAATGTAAGA	4604
Db	17773	CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	17714
Qy	4605	AAAAAATGTAAGAAAAAGTTTCAATTTCTTTGCCAGCAACGTTATTCAAAATTCCTG	4664
Db	17713	TGTGTTGGTCACTCTGTAGTCCAGTTTACTCAGGAGGCTGAGGTGGGAGGATCAGCTG	17654
Qy	4665	AGCCCTTTTACTTCCCAAAATTTCTGCACTTTCTGCCCGCTA-----CCATTAGG	4712
Db	17653	AGCCCGGAAGTTGAGGCTGCAGTGAAGTGTGATCGCAACCATTCACACTACAGCCTGGGTG	17594
Qy	4713	TGACAGCACTAGTCCCAAAATTTGATTAATGCAATTTCTGGAAAAAGACTAGGGCAAAAT	4772
Db	17593	ACAGAGCAAGACTCCGTTCTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA	17534

Query Match	4.18;	Score 226.8;	DB 9;	Length 32204;
-------------	-------	--------------	-------	---------------

Query Match	Score 226.8;	DB 9;	Length 32204;
4.18;	Score 226.8;	DB 9;	Length 32204;

[illegible]

Db	15305	ACGCCCACTAATTTTTTTTGTATTTTGGTAGACGGGGTTTTACTACGTTGGCCAGGC	15364
Qy	5490	TGGTCTCAAACTCCTGACCTCA	5511
Db	15365	TGGTCTCAAACTCCTGAACTCA	15386

RESULT 7

US-10-073-961-399

; Sequence 399, Application US/10073961

; Publication No. US20030077602A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA113C1

; CURRENT APPLICATION NUMBER: US/10/073,961

; CURRENT FILING DATE: 2002-03-05

; PRIOR APPLICATION NUMBER: 09/764,887

; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: 60/179,065

; PRIOR FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: 60/214,886

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 60/217,487

; PRIOR FILING DATE: 2000-07-11

; PRIOR APPLICATION NUMBER: 60/225,758

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/220,963

; PRIOR FILING DATE: 2000-07-26

; PRIOR APPLICATION NUMBER: 60/217,496

; PRIOR FILING DATE: 2000-07-11

; PRIOR APPLICATION NUMBER: 60/225,447

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/218,290

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 60/225,757

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/226,868

; PRIOR FILING DATE: 2000-08-22

; PRIOR APPLICATION NUMBER: 60/216,647

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: 60/225,267

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/216,880

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: 60/225,270

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/251,869

; PRIOR FILING DATE: 2000-12-08

; PRIOR APPLICATION NUMBER: 60/235,834

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: 60/234,274

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: 60/234,223

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: 60/228,924

; PRIOR FILING DATE: 2000-08-30

; PRIOR APPLICATION NUMBER: 60/224,518

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/236,369

; PRIOR FILING DATE: 2000-09-29

; PRIOR APPLICATION NUMBER: 60/224,519

; PRIOR FILING DATE: 2000-08-14

; PRIOR APPLICATION NUMBER: 60/220,964

; PRIOR FILING DATE: 2000-07-26

; PRIOR APPLICATION NUMBER: 60/241,809

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/249,299

; PRIOR FILING DATE: 2000-11-17

; PRIOR APPLICATION NUMBER: 60/236,327

```

, PRIOR FILING DATE: 2000-09-29
, PRIOR APPLICATION NUMBER: 60/241,785
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: 60/244,617
, PRIOR FILING DATE: 2000-11-01
, PRIOR APPLICATION NUMBER: 60/225,268
, PRIOR FILING DATE: 2000-08-14
, PRIOR APPLICATION NUMBER: 60/236,368
, PRIOR FILING DATE: 2000-09-29
, PRIOR APPLICATION NUMBER: 60/251,856
, PRIOR FILING DATE: 2000-12-08
, PRIOR APPLICATION NUMBER: 60/251,868
, PRIOR FILING DATE: 2000-12-08
, PRIOR APPLICATION NUMBER: 60/229,344
, PRIOR FILING DATE: 2000-09-01
, PRIOR APPLICATION NUMBER: 60/234,997
, PRIOR FILING DATE: 2000-09-25
, PRIOR APPLICATION NUMBER: 60/229,343
, PRIOR FILING DATE: 2000-09-01
, PRIOR APPLICATION NUMBER: 60/229,345
, PRIOR FILING DATE: 2000-09-01
, PRIOR APPLICATION NUMBER: 60/229,287
, PRIOR FILING DATE: 2000-09-01
, PRIOR APPLICATION NUMBER: 60/229,513
, PRIOR FILING DATE: 2000-09-05
, PRIOR APPLICATION NUMBER: 60/231,413
, PRIOR FILING DATE: 2000-09-08
, PRIOR APPLICATION NUMBER: 60/229,509
, PRIOR FILING DATE: 2000-09-05
, PRIOR APPLICATION NUMBER: 60/236,367
, PRIOR FILING DATE: 2000-09-29
, PRIOR APPLICATION NUMBER: 60/237,039
, PRIOR FILING DATE: 2000-10-02
, PRIOR APPLICATION NUMBER: 60/237,038
, PRIOR FILING DATE: 2000-10-02
, PRIOR APPLICATION NUMBER: 60/236,370
, PRIOR FILING DATE: 2000-09-29
, PRIOR APPLICATION NUMBER: 60/236,802
, PRIOR FILING DATE: 2000-10-02
, PRIOR APPLICATION NUMBER: 60/237,037
, PRIOR FILING DATE: 2000-10-02
, PRIOR APPLICATION NUMBER: 60/237,040
, PRIOR FILING DATE: 2000-10-02
, PRIOR APPLICATION NUMBER: 60/240,960
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: 60/239,935
, PRIOR FILING DATE: 2000-10-13
, PRIOR APPLICATION NUMBER: 60/239,937
, PRIOR FILING DATE: 2000-10-13
, PRIOR APPLICATION NUMBER: 60/241,787
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: 60/246,474
, PRIOR FILING DATE: 2000-11-08
, PRIOR APPLICATION NUMBER: 60/246,532
, PRIOR FILING DATE: 2000-11-08
, PRIOR APPLICATION NUMBER: 60/249,216
, PRIOR FILING DATE: 2000-11-17
, PRIOR APPLICATION NUMBER: 60/249,210
, PRIOR FILING DATE: 2000-11-17
, PRIOR APPLICATION NUMBER: 60/226,681
, PRIOR FILING DATE: 2000-08-22
, PRIOR APPLICATION NUMBER: 60/225,759
, PRIOR FILING DATE: 2000-08-14
, PRIOR APPLICATION NUMBER: 60/225,213
, PRIOR FILING DATE: 2000-08-14
, PRIOR APPLICATION NUMBER: 60/227,182
, PRIOR FILING DATE: 2000-08-22
, PRIOR APPLICATION NUMBER: 60/225,214
, PRIOR FILING DATE: 2000-08-14
, PRIOR APPLICATION NUMBER: 60/235,836
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: 60/230,438
, PRIOR FILING DATE: 2000-09-06

```

1	PRIOR APPLICATION NUMBER: 60/215,133
2	PRIOR FILING DATE: 2000-06-30
3	PRIOR APPLICATION NUMBER: 60/225,266
4	PRIOR FILING DATE: 2000-08-14
5	PRIOR APPLICATION NUMBER: 60/249,218
6	PRIOR FILING DATE: 2000-11-17
7	PRIOR APPLICATION NUMBER: 60/249,208
8	PRIOR FILING DATE: 2000-11-17
9	PRIOR APPLICATION NUMBER: 60/249,213
10	PRIOR FILING DATE: 2000-11-17
11	PRIOR APPLICATION NUMBER: 60/249,212
12	PRIOR FILING DATE: 2000-11-17
13	PRIOR APPLICATION NUMBER: 60/249,207
14	PRIOR FILING DATE: 2000-11-17
15	PRIOR APPLICATION NUMBER: 60/249,245
16	PRIOR FILING DATE: 2000-11-17
17	PRIOR APPLICATION NUMBER: 60/249,244
18	PRIOR FILING DATE: 2000-11-17
19	PRIOR APPLICATION NUMBER: 60/249,217
20	PRIOR FILING DATE: 2000-11-17
21	PRIOR APPLICATION NUMBER: 60/249,211
22	PRIOR FILING DATE: 2000-11-17
23	PRIOR APPLICATION NUMBER: 60/249,215
24	PRIOR FILING DATE: 2000-11-17
25	PRIOR APPLICATION NUMBER: 60/249,264
26	PRIOR FILING DATE: 2000-11-17
27	PRIOR APPLICATION NUMBER: 60/249,214
28	PRIOR FILING DATE: 2000-11-17
29	PRIOR APPLICATION NUMBER: 60/249,297
30	PRIOR FILING DATE: 2000-11-17
31	PRIOR APPLICATION NUMBER: 60/232,400
32	PRIOR FILING DATE: 2000-09-14
33	PRIOR APPLICATION NUMBER: 60/231,242
34	PRIOR FILING DATE: 2000-09-08
35	PRIOR APPLICATION NUMBER: 60/232,081
36	PRIOR FILING DATE: 2000-09-08
37	PRIOR APPLICATION NUMBER: 60/232,080
38	PRIOR FILING DATE: 2000-09-08
39	PRIOR APPLICATION NUMBER: 60/231,414
40	PRIOR FILING DATE: 2000-09-08
41	PRIOR APPLICATION NUMBER: 60/231,244
42	PRIOR FILING DATE: 2000-09-08
43	PRIOR APPLICATION NUMBER: 60/233,064
44	PRIOR FILING DATE: 2000-09-14
45	PRIOR APPLICATION NUMBER: 60/233,063
46	PRIOR FILING DATE: 2000-09-14
47	PRIOR APPLICATION NUMBER: 60/232,397
48	PRIOR FILING DATE: 2000-09-14
49	PRIOR APPLICATION NUMBER: 60/232,399
50	PRIOR FILING DATE: 2000-09-14
51	PRIOR APPLICATION NUMBER: 60/232,401
52	PRIOR FILING DATE: 2000-09-14
53	PRIOR APPLICATION NUMBER: 60/241,808
54	PRIOR FILING DATE: 2000-10-20
55	PRIOR APPLICATION NUMBER: 60/241,826
56	PRIOR FILING DATE: 2000-10-20
57	PRIOR APPLICATION NUMBER: 60/241,786
58	PRIOR FILING DATE: 2000-10-20
59	PRIOR APPLICATION NUMBER: 60/241,221
60	PRIOR FILING DATE: 2000-10-20
61	PRIOR APPLICATION NUMBER: 60/246,475
62	PRIOR FILING DATE: 2000-11-08
63	PRIOR APPLICATION NUMBER: 60/231,243
64	PRIOR FILING DATE: 2000-09-08

Query Match

Query Match	4.1%	Score 225.4;	DB 9;	Length 28588;
Best Local Similarity	48.3%	Pred. No. 1.1e-36;		

Matches 1154; Conservative 1; Mismatches 1142; Indels 94; Gaps 15;

Oy 3213 TTTATGTTTTTATGTTTTTGATTTTTAAAGACAATCTCACCTGTTCACCAAGGTGG 3272
||| || ||||| ||| ||||| | - ||| ||||| ||||| ||
Db 19521 TTTTTTTATTATTATTTTTTTTTTGATCGGAGTTTCACTCTGTTGCCCAGGCGGG 1958

Qy	3273	AGTGCAGTGGTGC	AAATCATAGCTTTCT	GCAGTCTTTG	AAACTCCTCGTGGCTCA	AGCAATCCTC	3332					
Db	19581	AGTGAAGTGGCGT	GTGACTCTCAGCTC	TGCAACCTCC	AGCTCCTCGTGGTTC	CAAGCGATTTCTC	19640					
Qy	3333	CTGCCCTGGCCCT	CCCAAAAGTGT	TGGGAT---	ACAGTCA	TGAGCCACTGCATCTG	GCCTAG	3389				
Db	19641	CTGCCCTGAGCT	CTCGTAGTGGGAT	TACAGT	TGCCCAACACAT	CCAGCTAAATTT	19700					
Qy	3390	GATCCATTTAGAT	TAAATATG	CACTTTTAAAT	TTTAAATATAT	TGG----CTAATTTTTT	3445					
Db	19701	TTTATATTTTAG	TAGAGAGGGG	TTTACCGTGT	TGACCAAGCTGGTCT	CAAACTCCTG	19760					
Qy	3446	ACCTTATGTAAT	-----GTG	TACTGGTAA	TAAATCTAG	TTTGGCTCA	AAAGTTTA	3498				
Db	19761	ACCTCAGGTGAT	CCACCTTGGCCT	CCCCAAAGT	TGGGATTTAC	AGGCATGAGCCAC	19820					
Qy	3499	AGTGGCTTTCCA	TAAAGCTTC-----	ATGTAC	GTGAGGGAGACAT	TTTAAAGTGA	AAACA	3552				
Db	19821	CGTGGCTTGCC	AAAAAATTTTTT	CTGATTTTAG	AGGAAAGTGT	TATAAAAATAG	TTATT	19880				
Qy	3553	GACAGCAGGTGT	GTGGCTCAGCGCT	GTAAATCCC	AGCACCTCTGG	AGGCTGAGGTGGGT	3612					
Db	19881	GCAGGCTGGC	ACGGTGGCTCAT	GGCTGTAA	TCCCAGCACAT	TGGAGGCTGAG	CGCAGGT	19940				
Qy	3613	GGATCGCTTG	AGCCCTGGAGTT	CAAGAC	AGCCTGAGCA	CAATGCA	AAAAACCCCTG	3672				
Db	19941	GGATCACTTG	AGCTTAGGAGTT	TGAGATCGG	CTCAGCAACAT	GSCAAAACCCCTG	CTTT	20000				
Qy	3673	ATAACAAA-----	AATTTAG	CCGGCATGGT	GCATGGC	GTGTGCTGCC	AGTACTA	3724				
Db	20001	ACCAAAATACA	AAACCAATTTAG	CAAGTGTG	CACTGGCAT	TGGCTGTGCTGCC	AGATCTC	20060				
Qy	3725	GGGGCTGAGG	CAGGAGAATCTTT	TGGAGCC	AGGAGGTCA	AGGCTGCACTG	AGCAGTGCT	3784				
Db	20061	AGAGGCTGAG	CAGGAGGATAC	CTTGAGGCT	CGGAGCGA	AGGTGCACTG	AGCTGACAT	20120				
Qy	3785	TGGCCACTGCA	CTCCAGCCTGGGT	GCAGCAG	ACCAGACCTTT	GCCTCA	AAAAAAATAAG	3844				
Db	20121	CGTGCCACT	TGCAATCAGCCT	TGGCAAC	AGAGTGAGAT	CCTGTCTCA	AAAAAACAAAAAC	20180				
Qy	3845	AAAAATTA	AAATTAATG	AAACAACT	ACAAGACG-----	TGTTGTCC	3888					
Db	20181	AAAAACGGT	TATAGCATAC	CAAGGCTTTAT	GTATATAG	ATAAGAAAT	GAGTTTTTTTTATT	20240				
Qy	3889	TAGATGAGCT	ACTTAGT	TAGGCTGAT	TTTTCGTTATTT	TAACCTTTTAA	AGTCAGGCTGTG	3948				
Db	20241	TTCACTGTC	ATCTTAATAT	ATATCATATTT	GCATAAAGCC	CTAGTGAATTTT	TAGTTTCA	20300				
Qy	3949	CACCTGCAC	TACATATTTAA	ATAATCAAT	TTCTCAAT	GTATATCC	ACACAAAGCTGGTAC	4008				
Db	20301	AGACTTAAT	TGGTTTTTTTTT	CTTTTTTAA	ATTTACATTT	TGAAATGAGG	CTTTTCATATAC	20360				
Qy	4009	GTCAATGTT	CATAGTAC-CT	TTTATTC	ACAAAACCCCA	AGTAGAGACT	TATCCAATATCC	4067				
Db	20361	TATACCTTT	TATATAC	GAATGTTAA	TTTAAATAG	CACTTTAC	AGTATAA	20420				
Qy	4068	ATCAACAAG	TGAACAAATA	AAAAATG	CTTATCC	ATGCAAT	TGGAATACCA	CCCTGC	4127			
Db	20421	ATTTTTCA	TAGATGCT	TAAACAC	CCATTTCC	TATTTAC	ATTTTAT	TCCAGTA	20480			
Qy	4128	AGTACA	AAAGGAAGC-----	TAC	TTGGGATGA	ATCCCA	AGTCACTAG	CCCTAAATGAA	4183			
Db	20481	TAGTTAA	AGATTTAAG	CCATTGTA	ACTTTGAC	AGGTAAATAT	ATAA	CACTATTAG	20540			
Qy	4184	AGAGTC	AGATGAAGG	AGGAGAT	TATGTAT	CCCATAG	CAAAAT-TCT	TAGAAAATGA	AGT	4242		
Db	20541	AATGTCT	CTCCAC	AAAGGAAC	TGAACTTAT	TGCTTGAT	AGTGTCT	TTTGGTTTATA	AAA	20600		
Qy	4243	AACTTAT	ATGATAC	AGAAAC	CAATCAG	GCAGC	ATAGGG	CTCAC	CTCTAAT	CCCG	4302	
Db	20601	AAGTAG	AAATGA	AGTATAT	TATCCT	TGGAGC	ATGTGAA	ATTCAT	TGAC	CCCTTTG	GCACAAAG	20660
Qy	4303	CAC	TTTTGAG	AGGCC	AGCTGGG	AGATTTG	TAGAACT	TCAGG	AGTTTCA	AGCCAG	CGCTGGGC	4362

Db	20661	GCTAAATAGAATATTAGTAGAATATTTCAGTAATAATAGACTTTGAACCTGCTGTTTTTCTC	20720
Qy	4363	AACACAGTGAACCTCCATTCTCCACAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGG	4422
Db	20721	TGTAFTGTTAAATAAATTTTTTCAAGAAATACAAAGGTTAGGCTATGAACAATTTATGATTT	20780
Qy	4423	TCCTGTGGGGGGAAGGACTGCAAGAGGGAAGA--AGCTCTGGTGGGGTGAGGGTG	4479
Db	20781	TATTAAATCATAGAAAGGACTTACTTCTCATATTTGAGAAGTTTATATCATCTGTGCT	20840
Qy	4480	GTGATTCAGGTTCTGTATCCCTGACTGTGGTAGCAGTTTGGGGTGTTCATATCCAAAATA	4539
Db	20841	TTTACTTAGTTTTCAGCATAATTGCTTCCCTCATTTTAAAGTAAATTTCTCTAAGCAT	20900
Qy	4540	TTCTGTAGAAATATGATCCTTAATGGGTGGAGTTTACTGTATCTAAATATATACCTCAATG	4599
Db	20901	TTCTATGCTTTGTTCTCTGCTTCGTATGATATATATTTTCCCTCCTAAGACTCAAC	20960
Qy	4600	TAAGAAAAAATAATGTGAAGAAAAAGTTTCAATTTCTTCTGCCAGCAAACTTATTCAAAT	4659
Db	20961	ATTGAGCTTTACACCTAGCTTCTGAGAGTCACTGTAATGGAAAGAGAGCCAGATAATT	21020
Qy	4660	TCCTGAGCCCTTTACTTCGCAAAATTCCTGCACT-TCTGCCCGGTACCAATTAGGTGACAG	4718
Db	21021	TCTTGGCCATACCTTGCCTTCAAACTTATCCAGGATGTGTATTTTGCAAGCTTCTGGAG	21080
Qy	4719	CACTAGCTCCCAAAATTTGGATAAATGCATTTCTGGAAAGACTAGGGACAAAA--T	4772
Db	21081	CAGTTTCCATCTCCTATTATGATGTTTACCTCTGAATCGAGCTTGAAGACTTACATGTCC	21140
Qy	4773	CGAGCATCACTTGTGCTTTTCATATCAACCAGCTGTACAGCTGTGTGTTGCTGCTGCAG	4832
Db	21141	GAATGGAAGAAGCATTTCAAAAACGGNAATGGCAGTTGCCAGTTCTTGGAAATCTAATCCTT	21200
Qy	4833	CTGCAATGGGACTCTTGATTTCTTTAAGAAACTTGGGTACCAGAGTAT-----	4883
Db	21201	GGGTAGAAAAGGTTATTTATCTCGTATGTTAATTTGATTTCTTAAGCAGACTTACTAGGA	21260
Qy	4884	TTCCCAAATGCTATCAAAATTTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAG	4943
Db	21261	TTTTCAGCAGTTATCTGAGACATCGTGTATTTATGTAACATTTATTTGTAAGTTAGGTGCTT	21320
Qy	4944	AAAACAAAGAGGAGAGAAATCAGTCATTTA-----TGTGGGAA	4981
Db	21321	CACGTATTTGTTTGTTCATTTATTTATGAGCACTGCCATGTCTCAGGCTCTGTGGTGG	21380
Qy	4982	CAACATAGCAAGATATTTAGATCAATTTTGACCTAGTTTAAAAAGCAGCAGATCAAAATC	5041
Db	21381	TCAGGTGGAAGAAGATAGACCTGGTCTCTACTGTGTATTTGCTCTAGTTTGTAGTGATG	21440
Qy	5042	ACACATGCANTCAGTATTAATCCAAATCATGTAATATGTGCCGTGTAGAAGACTAGAGGA	5101
Db	21441	CTTGGATAGTAAATAACAGTTTATGCAATTAGTTTCTGCTGGGTCTGTGATTTGA	21500
Qy	5102	ATAACACAAGAAATCTTAAACAGTCATTTGCAATTAGACACTTAAGCTCTAAATATTATTATTA	5161
Db	21501	AGGCTTGTGTATACCTATTCAAGATTTTAAACATTAGACTTTTCAGTTGATTTGCTTATA	21560
Qy	5162	GACACTATGATATTTGAGATTTTAAAAATCTTTTAATATTTTAAAAATTTAGAGCTCTCTA	5221
Db	21561	ATCTTGTCTGTCTGCTATTTCAGCATTTAGCTGCACTTTAATCTGATAGAGGTTTGTCCCA	21620
Qy	5222	TTTTTCCATAGTATTTCAAGTTTTCAGAAATGATCAAGTATTTACICTTTCTTTTTTTTTTT	5281
Db	21621	GTTGAGACTTGATATTTAGGTGCTAAAAATTTCCCTCCTGAGGTTCTCTCTCTCTTTTTTTT	21680
Qy	5282	TTTTTTTTTTTTTTGAGATGGAGTTTGGTCTTGTGGCCATGCTGGAGTGGAAATGGCATGA	5341
Db	21681	TTTTTTTTTTTTTTTGACACGGAGTTTTCACCTC---TTGCCACAGCTGGAGTGCATGATGCA	21737
Qy	5342	YCATAAGCTCACTGCAACCTCCACTCCTGGGTTTCAAGCAAGCTGTGCCCTCAGCCCTCCC	5401

Db 21738 TTTCAGCTCACTGCAACCTCTGCCTCCCGGTTCAAGTGATCTCTTGGCTCAGCCTCCC 21797
 Qy 5402 GGGTAGATGGATTACAGCGCCACCACCACTCGGCTAATGTTGTATTTTAGTAG 5461
 Db 21798 GAGTAGCTGGGATTACAGTCCCGCCGCCACCACTGATTTTGTATTTTAGTAA 21857
 Qy 5462 AGATGGGTTTTCACCATGTTGGCCAGGCTGGTCTCAACCTCCTGACCTCAG 5512
 Db 21858 AGACGGGGTTTTCACCATGTTGGCCAGTCTGGTCTCGAATCCTGACCTCAG 21908

RESULT 8

us-09-764-887-399
 : Sequence 399, Application US/09764887
 : Patent No. US2002042096A1
 : GENERAL INFORMATION:
 : APPLICANT: Rosen et al.
 : TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 : FILE REFERENCE: P113
 : CURRENT APPLICATION NUMBER: US/09/764,887
 : CURRENT FILING DATE: 2001-01-17
 : Prior application data removed - consult PALM or file wrapper
 : NUMBER OF SEQ ID NOS: 658
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 399
 : LENGTH: 28588
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 us-09-764-887-399

Query Match 4.1%; Score 225.4; DB 10; Length 28588;
 Best Local Similarity 48.3%; Pred. No. 1.1e-36;
 Matches 1154; Conservative 1; Mismatches 1142; Indels 94; Gaps 15;

Qy 3213 TTTATGTTTTTATGTTTTTAAAGACAATCTCACCTGTTACCCAGGCTGG 3272
 Db 19521 TTTTATTTATTAATTTTATTTTGTAGATGGATTCTACTCTGTTCCCGAGCGGG 19580
 Qy 3273 AGTCAGTGTGCAATCATAGCTTCTGAGTCTTGAATCTCGGCTCAAGCAATCCCTC 3332
 Db 19581 AGTGAAGTGGGTTGCTGCACTGCACTGCAACCTCCACCTCTGGGTTCAAGCGATTCTC 19640
 Qy 3333 CTGCCTTGGCCTCCCAAGTGTGGGAT---ACAGTCATGAGCCACTGCATCTGGCCTAG 3389
 Db 19641 CTGCCTCAGCTCTCAGTAGCTGGGATTACAGGTGCCACCACTGAGCCAGCTAATTT 19700
 Qy 3390 GATCCATTAGATTAATAATGCAATTTTAAATTTTAAATAATATG---CTAATTTTT 3445
 Db 19701 TTTATATTTTATGATAGAGACGGGTTTCACCGTGTGACACAGCTGGTCTCAAACTCTG 19760
 Qy 3446 ACCTTATGTAAT-----GTGTATCTGGTAAATAATCTAGTTTCTGCTGCCTAAAGTTTA 3498
 Db 19761 ACCTCAGGTGATCCACCTGCTGGCTCCCAAGTGTGGGATTACAGGCATGAGCCAC 19820
 Qy 3499 AAGTGCTTTCCAAATAGCTTC-----ATGTACGTGAGGGGAGACATTTAAAGTGAACA 3552
 Db 19821 CGTGCCTGCCAAAAAATTTTTTCTGATTTTAGAGGAAGTGTATAAAATAGTATT 19880
 Qy 3553 GACAGCCAGTGTGGTCTGAGCTGTAATCCAGCACTCTGGAGGCTGAGGTGGT 3612
 Db 19881 GCAGGCTGGGACGGTGGCTGCTGCTGTAATCCAGCAATTTGGAGGCTGAGCGAGT 19940
 Qy 3613 GGATGGCTTGAGCCCTCGAGTTTCAAGACAGCTGAGCAACATGGCAACCTCTTCT 3672
 Db 19941 GGATCATTGAGCTTAGGNTTTGATGCGGCTGAGCAATGGCAAAACCTCTCTTT 20000
 Qy 3673 ATAACAAA-----AATTAGCGGGCATGGTGCATGTGCTGTGGTCCAGCTACTA 3724
 Db 20001 ACCAAAAATACAAACCAATTTAGCCAGGTGCAGTGGCATGTGCTGTGTCAGATACTC 20060
 Qy 3725 GGGGCTGAGGCGAGAGATCTTTGGAGCCAGGAGTCAAGGCTGACAGTGCCT 3784
 Db 20061 AGAGCTGAGGCGAGGATTACCTGAGCCTGGGAGCGAAGGTTGCGAGTGCAGT 20120

Qy 3785 TGCCCACTGCACCTCAGCCTGGGTGACAGCAGCAGACCTTGCCTCAAAAAAATAAGAAG 3844
 Db 20121 CGTGCCACTGCAATCCAGCCTGGGCAACAGAGTGAGATCCTGTCTCAAAAAAATAAGAAG 20180
 Qy 3845 AAAAATTTAAATTAATGGAACAACTACAAAGAGC-----TCTTCTCC 3888
 Db 20181 AAAAAGGTTATAGCATAGCAAGGCTTTATGATATGATATAAGAAATCAGTTTTTTTATT 20240
 Qy 3889 TAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTACAGGCTGT 3948
 Db 20241 TTCATGTCCATCTTAATTAATCAATAATTTGATAAGAGCCCTAGTGAATTTTGTAGTTCA 20300
 Qy 3949 CAGCTGCACATACATTAATAAATATCAATTTCAATGTATATCCACACAAAGACTGTAC 4008
 Db 20301 AGACTTAATGGTTTTTTTTTCTTTTTTAAATTACATTTGAAATGAGGCTTTCAATATAC 20360
 Qy 4009 GTGAATGTTTCATAGTAG-CTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCC 4067
 Db 20361 TATACCTTTTATATACGAATGATGTTAATTTTAAATAGCACTTACAGTATAAATTTAC 20420
 Qy 4068 ATCAACAGTGAACAAATAAACAATGCTGCTATATCCATGCAATGGAATACCACTGTC 4127
 Db 20421 ATTTTCAATAGATGCTTAACACCACTTTCCCTATTTACATTTGCTTAGTTTATCTCCAGTA 20480
 Qy 4128 AGTACAAAGGAAGAGC---TACTTGGGGATCAATCCCAAGTCAATGAGCTTAAATGAA 4183
 Db 20481 TAGTTAAAGATTAAAGCCATTGTAACCTTGACAGGTAAATATAAATGATTTAGGAATAA 20540
 Qy 4184 AGAGTCAGCATGAAGGAGAGATAATGTATGCCATACGAAAT-TCTAGAAAATGAAGT 4242
 Db 20541 AATGTCTCCACAGGGAACCTGAAACTTATTCCTTGTAAAGTGTCTTTGGTTTATAAA 20600
 Qy 4243 AACTTATAGTTACAGAAAGCAATCAGGCGAGCCTAGAGGCTCAGACCTGTAAATCCAG 4302
 Db 20601 AAGTAGAAATGAAGTATATATCCTGGAGGCTGTGAATCATTTGACCGCTTTGACACAAAG 20660
 Qy 4303 CACTTTGAGAGCCAGCTGGGAAGATTGCTAGAACTCAGAGTTCAAGACCAAGCTGGGC 4362
 Db 20661 GCTAAATAGATATTTAGTAGAATATTCAGTAATAATAGGACTTGAACCTGCTGTTTCTC 20720
 Qy 4363 AACACAGTGAACCTCATTCTCCACAAATGGGAAAAAAGAAAGCAATCAGTGGTTG 4422
 Db 20721 TGTATGTTTAAATAATTTTTCAAAGAAATACAAAGTTAGGCTATGAACATTTATGATTT 20780
 Qy 4423 TCTGTGGGAGGAGGAGCACTGCAAGAGGGAAGA--AGCTCTGGTGGGTGAGGCTG 4479
 Db 20781 TATTAAATCATAGAAAGGACTACTTATCTTCATATTGAGAAAGTTTATCATTTGTGCT 20840
 Qy 4480 GTGATTCAGGTTCTGTATCTGCTGACTGTGTAGCAGTTTGGGTGTTTACATCCAAAAATA 4539
 Db 20841 TTTTACTTACTTTTACGATATTTGCTTCCCTCAATTTTAAAGTAAATTTCTCTAAGCAT 20900
 Qy 4540 TTTCTGAGAAATATGATCTTAAATGGTGGAGTTTACTGTATGTAAATATACCTCAATG 4599
 Db 20901 TTTCTATGCTTTGCTTCTGCTTCTGATTTGATATATTTTCTCTCTAGAACTCAAC 20960
 Qy 4600 TAAAGAAAAAATAATGTGAAGAAAAAGTTTCAATTTCTGCGCAGCAAAAGTTTATTCAAT 4659
 Db 20961 ATTGAGCTTTACACCTAGCTCTGAGAGTCACTGTAAATGGAAGAGAGAGGAGGAGT 21020
 Qy 4660 TCTGTAGCCCTTTACTTCGCAAAATTTCTCTGCACT-TCTGCCCCGTACCATTAGGTGAC 4718
 Db 21021 TCTTGGCCATACCTTGCCTTCAAACTTATCCAGGATGTGTCAATTTTTCAGCTCTTGCAG 21080
 Qy 4719 CACTAGCTCCCAAAATGATTAATGCAATTTCTGGAAGAACTAGGGCAAAA-----T 4772
 Db 21081 CAGTTCCATCTCTCTATGATTTGTTACCTCTGCAATCGAGGCTCTGAAGACTCTACATGTC 21140
 Qy 4773 CCAGCATCTCTGCTTTTCTATCATCAACACCGCTGTACAGCTTGTGTTGCTGCTGAC 4832
 Db 21141 GAATGGAAAGCAATTTCAAAACCGAATGGCAGTTGGCCAGTTTCCCTGGAATCTAATCCTT 21200

Db 23618 GGAAAAAATAATTAATAAATAATGCGGAGTGTGCTGCTCACACCTGTATATCCAGC 23559
 QY 3591 ACTCTGGAGGCTGAGGTGGGTGATCGCTTGAGCCCTGGAGTTCAGACCCAGCCTGAGC 3650
 Db 23558 ACTCTGGAGGCTGAGGTGGGTGATCGCTTGAGCCCTGGAGTTCAGACCCAGCCTGAGC 23499
 QY 3651 AACATGGCAAAACCTGTTCT-----ATAACAAAAATAGCCGGCATGGTGGCATGT 3704
 Db 23498 AACATGGTGAACCTGCTCTCTACTATAAATAACAAAAATAGCTGGCGGTGTGTGGC 23439
 QY 3705 GCGTGTGCTCCAGCTACT-AGGGGCTGAGCGAGGAGATCTTTGGACCCAGGAGGTC 3763
 Db 23438 ATCTGTAAATCCAGCTACTCAGGAGCTGAGCGAGGAGATGCTTTGAGCCAGGAGGCA 23379
 QY 3764 AAGGCTGCACTGAGCAGTCTTGGCCACTCCAGCCTGAGCTGGGTGACAGGACCAAGC 3823
 Db 23378 GAGGTGCACTGAGTGTGATGATGCTGAGCTGAGCTGAGCTGAGCTGAGCT 23319
 QY 3824 TTGCTCTCAAAAAATAAGAGAAATAATAAATAAATAAATAAATAAATAAATAAATAA 3883
 Db 23318 CTGTCTCAAAAAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 23259
 QY 3884 TGTCTCTAGATGAGCTACTTAGTGTAGCTCA-----TATTTTGGTATTTAAAGT 3938
 Db 23258 TGGCTCACACCTGTATACAGCACTATGGAAGCTGAGGCGAGGAGGATTTGCTTAAGCCCT 23199
 QY 3939 CAGGCTGTCTACCTGCACTACATTAATAAATAATCAATTTCTCAATGTATATCCACAA 3998
 Db 23198 AGAGTTCAAGACCACTGAGCACTGAGCACTGAGCACTGAGCACTGAGCACTGAGCACT 23139
 QY 3999 AGACTGGTACGTGAATGTCATAGTACCTTTTATTCACAAACCCCAAGTAGAGACTATC 4058
 Db 23138 AAATCAATAAATAATGGCAGTTTCTCAGTTTCTGCTGCTGGGTCTATGCTGCTACTGTC 23079
 QY 4059 CAAATATCCATCAACAAGTGAACAAATAAACAATAATGCTATATCCATGCAATGAATA 4118
 Db 23078 TCTTCAGATCACTGATGCGCAGTTCACCTTCAAGACTGGAAGCAGCAGCAGCAAGGA 23019
 QY 4119 CCACCTCGAGTACAAAGGAAGAGTACTTGGGATGAATCCCAAGTCATGAGCCTAA 4178
 Db 23018 TGGGATTTTCGAGGCGCAACACTCCAGGTATTTTAAATTTTAAATTTAAATTTATAC 22959
 QY 4179 ATGAAGAGTACAGATGAAGGAGGAGATATGATGCAATACGAATTTCTAGAAATGA 4238
 Db 22958 AGGGCCAGCAGCGTGTACACCTGTAATTCGGGTACTTTGGGAGAACATGCAAAA 22899
 QY 4239 AAGTAACCTATAGTTACAGAAAGCAATCAGGCGAGGATAGAGCTCACACCTGTAATC 4298
 Db 22898 CCCTGTCTCTACTAAAAATAACAAAAATCAGGCTGGCGCGGTAGCTCATGCTGTAATC 22839
 QY 4299 CCAGCACTTTGAGAGCCACGTG-GGAAGATTGCTAGAACCTCAGGAGTTCAAGACAGCC 4357
 Db 22838 CCAGCACTTTGAGAGCCACGTGAGTACCTGAGATCAGGAGTCTGAGCCAGCC 22779
 QY 4358 TGGCAACACAGTGAACCTCCATCTCCACAAAAATGGAAAAAAGAAAGCAATCAGT 4417
 Db 22778 TGGCAACATGGCAACCCCACTCTACTAAAAATAACAAAAATAGCCGGCATGGTGT 22719
 QY 4418 GGTGTCTCTGT-----GGGAGGGGAGGAGTGAAGAGGAGGAGTCTGGTGGGTG 4473
 Db 22718 TGTACACCTTTAATCCCACTACTAAGGAGGCTGAGGCAAGAGATTTCTTGAACCTGGG 22659
 QY 4474 AGGGTGGTGAATCAGTCTGTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4533
 Db 22658 AAGCGGTGGTGCAGCAAGTCAAGATTGTGTGCCATTCGAATCCAGCCTGGGTGACAGCA 22599
 QY 4534 AAAATATTCGTAGAATTTATGATCTTAATGGTGGAGTTTACTGTATGTAATTTATACC 4593
 Db 22598 AGACTCCGCTCATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 22539
 QY 4594 TCAATGTAAAGAAAAAATAATGTGAAGAAAAATTTCAATTTCTTTCCGCAACAGCTTAT 4653
 Db 22538 AATTAGCTGGGTGTGTGTGCTCTACCTGTAGTCCCACTTACTCAGGAGGCTGAGGTGGG 22479

RESULT 11

US-10-072-349-328
 ; Sequence 328, Application US/10072349
 ; Publication No. US20030054420A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P1110C1
 ; CURRENT APPLICATION NUMBER: US/10/072.349
 ; CURRENT FILING DATE: 2002-02-11
 ; Prior Application removed - See file Wrapper or Palm

QY 4654 TCAATTTCTGAGCCCTTTACTTTCGCAAAATTTCTCTGCACCTTTCTGCCCG----- 4702
 Db 22478 AGGATCACCTGAGCCCGGGAAGTTGAGGCTGCACTGAGCTGTGATCGCACCATTTGCACTA 22419
 QY 4703 -TACCATTTAGGTGACAGCACTAGCTCCACAAATTTGGATAAATGCAATTTCTGGAAGAGCT 4761
 Db 22418 CAGCCTGGGTGACAGAGCAAGACTCGCTCTTAAAAAATAAATAAATAAATAAATAAATAA 22359
 QY 4762 AGGCAAAATCCAGGCATCAGCTTGTGCTT-TCATATCAACCAAGCTGTACACCTTGTGT 4820
 Db 22358 GTGATCACAGCTCAGCTGTAACTTTGAATCCTTTACTTCAAGCAGTCTTACACAGCCTAC 22299
 QY 4821 TGTGTCTCTCAGCTGCAATGGGAGCTCTTGATTTCTTAAAGAAATCTGGTTTACCAGAG 4880
 Db 22298 AGGTGTGTCACCCCTGCTTAATTTTAAAAATTTTGTAGAGATGGGTGATGCTA 22239
 QY 4881 TATTTCCCAAAATGCTATTTCAAAATAGTCTTATGATATGCAAGACACTGTGCTAGGAGC 4940
 Db 22238 TGTGGCCAGGCTGCTCTCAAACTCCTGCCTCAAGTGAGCCTCTGCCTTGGCCTCCCT 22179
 QY 4941 CAGAAACAAAGAGGAGGAGAAATCAGTCAATATGTGGGNAACACATAGCAAGATATTTA 5000
 Db 22178 AAGTGTGGATTTATAGTGGGAGC-----CACTGCAATCG 22143
 QY 5001 GATCATTTTCTAGTGTAAAGAGCAGAGAGTACAAAAATCACACATGCAATCAGTATAA 5060
 Db 22142 GTTGCATTTCTTATTAATTAATACATCTTCTTATCATATCAATATACATATCCTCAT 22083
 QY 5061 TCCAAATCATGTAATATATGCTGCTGTAGAAAGACTAGAGAAATAAAC-ACAAGAAATCTTA 5119
 Db 22082 ATTAAGTGTGATAGAACAGCAGTGTGGGTGTAGTTTATTTAGCTATTTCTTTTTTTC 22023
 QY 5120 ACAGTCATTTGATAGACACTAAGTCAATATTTATTTATTTAGACACTATGATTTGAG 5179
 Db 22022 ACGAAATTTATGTTTGTACAGTTTTTGTCTATTACCAAGAGTGTGACAGTGAACATTT 21963
 QY 5180 ATTTAAAAATCTTTAATAATTTTAAATTTAGAGCTCTTCTATTTTCCATAGTATTCAA 5239
 Db 21962 CTGTGACACACCTTTTATACATATATGCTATTATTTGAGATAGAGAAGT---GG 21907
 QY 5240 GTTTCACATGATCAAGTATTACTCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGATG 5299
 Db 21906 GATTCGTGAGTGAAGATATGATTTGTTTTTTGGGTTTTTTTTTTTTTTTTTTTTTTGAGATG 21847
 QY 5300 GAGTTTGTCTGTGTGCCCCTGCTGGAGTGGCAATGGCATGAYCATAGCTCACTGCAACC 5359
 Db 21846 GAGTTTGTCTGTGTGCCCAGCTGGAGTAAATGGCGTGA-CITGGCTCACTGNAACC 21788
 QY 5360 TCCACCTCTGTTGTTCAAGCAAGCTGTGCGCTCAGCCTCCCGGGTGTAGTGGATTTACAG 5419
 Db 21787 TCTGCTCCCTAGTTCAAGCAATTTCTCCTGCTCAGCCTCCCAAGTAGTGGGATTTACAA 21728
 QY 5420 GGGCCCAACACACACTCGGCTAA--TGTTTCTATTATTAGTAGAGATGGGTTTCACCA 5477
 Db 21727 GCACCCCAACACACCCAGCTAAATTTTTTGTATTTTTGGTAGAGAGCGGTTTTACTTA 21668
 QY 5478 TGTGGCCAGGCTGCTCAAACTCCTGACCTCA 5511
 Db 21667 CGTTGGCCAGGCTGCTCCAACTCCTGAACCTCA 21634

; NUMBER OF SEQ ID NOS: 334									
; SOFTWARE: PatentIn Ver. 3.1									
; SEQ ID NO 328									
; LENGTH: 32152									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-10-072-349-328									
Query Match									
Best Local Similarity 4.0%; Score 221.2; DB 9; Length 32152;									
Matches 984; Conservative 0; Mismatches 968; Indels 62; Gaps 12;									
QY	3531	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGGTGGCTCACGCCCTGTATCCAGC	3590						
DB	8535	CGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	8594						
QY	3591	ACTCTGGGAGGTGAGGTGGGTGGATCGCTTTCAGCCCTGGAGTTCAAGACCGCTGAGC	3650						
DB	8595	ACTCTGGGAGGCAAGACCGGTGGATCACCTGAGGTTCAGAGCTTCAAGACCGCTGGCC	8654						
QY	3651	AACATGGCAAAACCCCTGTTTCT-----ATAACAAAAATTAGCCGGCATGGTGCACTG	3704						
DB	8655	AACATGGTGAACCTCGTCTCTACTAAAAATTACAAAAATTAGTGGCGTGGTGGGC	8714						
QY	3705	GCCTGTGGTCCAGCTACT-AGGGGGCTGAGCAGAGAAATCTTTGGAGCCAGGAGTGC	3763						
DB	8715	ATCTGTAAATCCAGCTACTCAGGAGCTGAGCAGGAGAAATGGCTTGAGCCAGGAGGCA	8774						
QY	3764	AGGCTGCAGTGCAGTGGCTTGGCCACTGCAGCTCCAGCTGGGTGGTGACAGACCGACC	3823						
DB	8775	GAGGTGTCAGTAAGCTGAGATGCCACTGCAGCTGCAGCTCCAGCTTGGAGACAGAGTGAGCT	8834						
QY	3824	TTGGCTCAAAAAATAAGAGAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	3883						
DB	8835	CTGTCTCAAAAAACAACAAACAACAACAACAACAACAACAACAACAACAACAACAACA	8894						
QY	3884	TGCTCTAGTAGCTACTTAGTTAGGCTGA-----TAATTTGGTATTTAACTTTTAAAGT	3938						
DB	8895	TGGCTCACACCTGTAAATACAGCAGCTATGGAAGCTGAGCAGGAGGATTCCTTAAGCCCT	8954						
QY	3939	CAGGCTCTGCAGCTGCAGTACATTAATAAATAATCAATTCATATGTATATCCACAA	3998						
DB	8955	AGAGTTCAGACAGCAGCTGGGCAATGTAGTGAGACCCCCCTCTCTACAGGAACAATA	9014						
QY	3999	AGACTGGTACGTGAATGTTTCATAGTACCTTTTACAAAAACCCCAAGTAGAGACTATC	4058						
DB	9015	AAATACAATAAAATGGCAGTTCTCAGTTTGTGCTGGGTGCTATGTGCTCAGTCTGTC	9074						
QY	4059	CAAAATCCATCAACAAGTGAACAAATAAACAATAATGTCTATATCCATGCAATGGAATA	4118						
DB	9075	TCTTCAGAAATCACTGATGCCAGGTTCACCTTCAAGACTTGAAGCAGCAGCACAGA	9134						
QY	4119	CCACCTGCAGTACAAGAGAGAGCTACTTGGGATGAATCCCAAGTCATGACCTAA	4178						
DB	9135	TGGGATTTTCAGAGCCAGAACCTCCAGGTATTTTAAATTTTAAATTTTAAATTTATAC	9194						
QY	4179	ATGAAGAGCTCAGACATGAAGGAGAGATAATGTATGCCATACGAAATTCAGAAATGA	4238						
DB	9195	AGGGCCAGCAGCGTGGCTCACACCTGTAATTCGGTACTTTGGGAGAACATPGGCAAAA	9254						
QY	4239	AGATACTTATAGTTACAGAAAGCAAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATC	4298						
DB	9255	CCCTGTCTCTACTAAAAATACAAAAAATCAGGCTGGCGGGTAGCTCATGCTGTATTC	9314						
QY	4299	CCAGCACTTTGAGGCCACGGT-GGAAGATGTCTAGAACTCAGGAGTTCAAGACCGACC	4357						
DB	9315	CCAGCACTTTGGGAGGCCAAGCAGGTGGATCACCTGAGATCAGGAGTCTGAGCCCGACC	9374						
QY	4358	TGGGCAACACTGAACTCCATCTCCACAAAAATGGGAAAAAAGAAAGCAATCACT	4417						
DB	9375	TGGCCAAACATGGCGAAACCCCTCTCTACTAAAAAATACAAAAAATAGCCGGCGAT	9434						
QY	4418	GGTTGTCTGT-----GGGAGGGGAGGACTGCAAGAGGGGAAGAGCTCTGGTGGGTG	4473						

DB	9435	TGTACACCTTTAATCCAGCTACTAAGAGGAGCTGAGCAAGAGAAATGCTTGAACCTGGG	9494						
QY	4474	AGGGTGGTGATTCAGGTTCTGTATCCTGACTGTGGTAGCAGTTGGGGTGTTTACATCCA	4533						
DB	9495	AAGCGTGGTTGCAGCAAGTCAAGATTTGTGTGCCATTTGCAATCCAGCCTGGGTGACAGCA	9554						
QY	4534	AAAAATTCGTAGAATTAATGCACTCTTAAATGGGTGGAGTTTACTGTATGTAATATATACC	4593						
DB	9555	AGNCTCCGTCTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	9614						
QY	4594	TCAATGTAAAGAAAAATAATGTGTAAAGAAAGTTTCAATTCCTCTGCCAGCAACCGTAT	4653						
DB	9615	AATTAGCTGGGTGTGGTGTGCTCACCTGTAGTCCAGTTACTCAGGAGCTCAGGTGGG	9674						
QY	4654	TCAATTCCTGAGCCCTTTACTTCGCAAAATTCCTGCACCTCTCTGCCCG-----	4702						
DB	9675	AGATCACCTGAGCCCGGGAAGTTGAGCTGCACTGAGCTGTGATCGCACCATTTGCACATA	9734						
QY	4703	-TACCAATTAGGTGACAGCACTAGCTCCACAAATTTGGATAAATGCAATTTCTGGAAGAGCT	4761						
DB	9735	CAGCCTGGGTGACAGAGCAAGACTCCGCTCTTAAAAAATAAATTTATTTTTTAGGGATAT	9794						
QY	4762	AGGACAAAAATCCAGGCATCATCTGTGCTT-TCATATCAACACAGCTGTACAGCTGTGT	4820						
DB	9795	GTGATCACAGCTCAGCTGTAACCTTGAACCTGCTTACTTCAAGCAGCTCTACCACAGCCTAC	9854						
QY	4821	TGCTGTCTGCAGCTCAATGGGGACTCTTGATTTCTTTAAGGAACCTTGGTTACACAGAG	4880						
DB	9855	AGGTGTGTCGCCCCCTGGCTAAATTTTTTAAATTTTTTGTAGAGATGGGCTCATGCTA	9914						
QY	4881	TATTTCCACAAATGCTATTCAAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGC	4940						
DB	9915	TGTTGCCAGGCTGCTCAACTCTGGCTCAAGTGAGCTCTCTGGCTTGGCTCCCT	9974						
QY	4941	CAGAAAAAAGAGGAGAGAAATCAGTCAATTAATGTGGGAACAACATAGCAAGATTTTA	5000						
DB	9975	AGTGCTGGGATTATAGTGGGAGC-----CACTGCAATTCG 10010							
QY	5001	GATCATTTTTCAGTGTATTAATAAGACAGCAGAGTACAAAAATCACACATGCAATCAATATA	5060						
DB	10011	GTTCGATTTCTTTTATTTAATTTACATCTTGATCTTTTCATATCAATATACATATCTCAT	10070						
QY	5061	TCCAAATCATGTAATATGTGCTGTAGAAAGACTAGAGGAATAAAC-ACAAGAATCTTA	5119						
DB	10071	ATTAACTGTGCATAGACAGCATAGTGGGTGTAGTTTATTTAGCTATCTTTTTTTC	10130						
QY	5120	ACAGTCATTTGTCATTAGACACTAAGTCAATTAATTAATTAATAGACACTATGATATTTGAG	5179						
DB	10131	ACGGAATTTATGATTTGTTACAGTTTTCCTGCTATTACCAAGAGTGTGACAGTGAACATT	10190						
QY	5180	ATTTAAAAATCTTTAATATTTTAAATTTTAGAGCTCTCTCTATTTTCCATAGTATTTCAA	5239						
DB	10191	CTTTGACACACCTCTTTATACATACATATGTGTTATTTTGAAGATAGAGAAGT-----	10246						
QY	5240	GTTTGACAAATGATCAAGTATTACTCTCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTT	5299						
DB	10247	GATTGCTGAGCTGAAAGATATGCAATTTGTTTTTTTGGGTTTTTTTTTTTTTTTTTTT	10306						
QY	5300	GAGTTTGGTCTTTGTTGCCCATGCTGGAGTGAAGTGGCATGAYCATAGCTCACTGCAACC	5359						
DB	10307	GAGTTTTCGCTCTTTGTTGCCAGGCTGGAGTAAATGGCGTGA-CTTGGCTCACTGAAACC	10365						
QY	5360	TCCACCTCTCTGGGTTCAAGCAAGCTGTGCCCTCAGCTCCCGGGTAGATGGGATTTACAG	5419						
DB	10366	TCTGCCTCCTAGTTTCAAGCAATTCCTCCTGCCCTCAGCTCCCAAGTAGCTGGGATTTCAA	10425						
QY	5420	GGGCCCCACACACCTCGCTAA--TGTGTTGTTATTTTATAGTAGAGTGGGTTCACCA	5477						
DB	10426	GCACCCACACACCGCCAGCTAATTTTTTGTATTTTGTGTAGACGGGGTTTTACTA	10485						
QY	5478	TGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511							

QY	4239	AAGTAACTTATAGTTTACAGAAAGCAAAATCAGGCGCAGCATAGAGCTCACACCTGTAATC	4299
Db	9255	CCCTGTCTACTAAAAATACAAAAATCAGCTGGCGCGGTAGCTCATGCTGTAATC	9314
QY	4299	CCAGCACTTTGAGAGGCCACGTG-GGAAGATTGCTAGAACTCAGAGCTTCAAGACACGCC	4357
Db	9315	CCAGCACTTTGGGAGGCCAGGAGGTGGATCACCCTGAGATCAGAGAGTCTGAGCCACGCC	9374
QY	4358	TGGGCAACACAGTGAATCCATTCTCCACAAAAATGGGAAAAAAGAAAGCAAAATCAGT	4417
Db	9375	TGGCAACATAGGGGAAACCCCATCTCTACTAAAAATACAAAAATAGCCGGCATGGTTG	9434
QY	4418	GGTTGTCCTGT- - - -GGGGAGGGAAGGACTGCAAGAGGGAAGACTCTGGTGGGGTG	4473
Db	9435	TGTACACCTTTAATCCCAGCTACTAAGGAGGCTGAGGCACAGAGAAATTCCTTGAACCTGGG	9494
QY	4474	AGGGTGGTGATTCAGGTTCCTGTATCCCTGACTGTGCTAGCACTTTGGGGTGTTACATCCA	4533
Db	9495	AAGCGGTGGTTGCAGCAAGTCAAGATTGTGTGCCATTGCAATCCAGCCTGGGTGACAGCA	9554
QY	4534	AAATATTGCTAGAAATATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATATATACC	4593
Db	9555	AGACTCCGTCTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	9614
QY	4594	TCAATGTAAGAAAAAATAATGTCTGAAGAAAAGTTTCAATTTCTCTTCCGCCAAACGTTAT	4653
Db	9615	AATTAGCTGGTGTGTGGTGTCTACCTGTAGTCCCAAGTTACTCAGGAGGCTGAGGTGGG	9674
QY	4654	TCAAAATCCTGAGCCCTTTACTTTCCGCAAAATCTCTGCACATTTCTGCCCCG- - - -	4702
Db	9675	AGGATCACCTGAGCCCGGGAAGTTGAGGCTGCACTGAGCTGTGATCGCACCATTTGCACTA	9734
QY	4703	-TACCATTAGGTGACAGCACTAGCTCCACAAATTTGGATAAATGCATTTCTGGAAAGACT	4761
Db	9735	CAGCCTGGGTGACAGAGCAAGATCCGCTCTTAAAAAATAAATTTATTTTATAGGATAT	9794
QY	4762	AGGGCAAAATCCAGGCATCAGTTTGCTT-TCATATCAACCAACGCTGTACAGCTTCTGT	4820
Db	9795	GTGATCACAGCTCACTGTAACTTTGAACCTCTTACTTCAAGCAGTCTCTACCACAGCCTAC	9854
QY	4821	TGCTGCTGCACTGCAATGGGACTCTTGATTTCTTTAAGGAACCTTGGGTACCAGAG	4880
Db	9855	AGGTGTGTCCACCCCTGGCTAAATTTTAAAAATTTTGTAGAGATGGGTCATGCTA	9914
QY	4881	TATTTCCAAATGCTATTCAAAATAGTGCTTATGATATGCAAGACACTGTGCTAGGAGC	4940
Db	9915	TGTTGCCAGGCTGCTCTCAAACTCCTGGCCTCAAGTGAGCCTCTGCTTGGCCTCCCT	9974
QY	4941	CAGAAACAAAGAGAGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTA	5000
Db	9975	AAGTCTGGGATTAATAGTGGGAGC- - - - -CACTGCATTCG	10010
QY	5001	GATCATTTTGTACTAGTTAAAAAAGCAGCAGAGTACAAAAATCACATATGCAATCAGTATAA	5060
Db	10011	GTTCGATTTCTTTATTTAAATACATCTTGATCTTTTCATATCAATATACATATCCCTCAT	10070
QY	5061	TCCAAATCATGTAATATGTGCTCTAGAAAAGACTAGAGGAATAAAC-ACAAAGAACTTTA	5119
Db	10071	ATTAACCTTGCATAGAACAGCATAGTGTGGGTAGTTTATTTAGCTATTTCTTTTTTTC	10130
QY	5120	ACAGTCATTGTCTAGACACTAAGTCTAAATATTATTATTAGACACTATGATATTTGAG	5179
Db	10131	ACGGAAATATGATTGTTTACAGTTTTTGTCTATTACCAAGAGTGTGACAGTGAACATT	10190
QY	5180	ATTTAAAAAATCTTTAAATATTTTAAAAATTTAGAGCTCTTCTATTTTCCATAGATTCAA	5239
Db	10191	CTTGTACACACCTCTTTATACATACATATGTGTATTATTGGAGATAGAGAAGT- - - -GG	10246
QY	5240	GTTTTGACATGATCAAGTATTACTCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGATG	5299
Db	10247	GATTGCTGAGCTGAAGAGATGCAATTCGTTTTTTTGGGTTTTTTTTTTTTTTTTTTTTCAGAT	10306

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:09:48 ; Search time 226.945 seconds
(without alignments)
6475.074 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagacgaatacataggtcca.....tcaaaactctgacctcagag 5514

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+L2p.model -DEV=xlp
-Q=Cyn2.1/USPto_spool/US09966880/runat_14062003_175523_10297/app_query.fasta_1.9493
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880.BCGN_1_1.632 &runat_14062003_175523_10297 -NCPU=6 -ICPU=3
-NO_MAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	345.5	3.6	361	23	ABG68738 Human prostate spe
2	334	3.5	110	21	AAB43292 Human ORFX ORF3056
3	332	3.4	202	22	AAE11962 Human c-myb-relate
4	311	3.2	105	22	ABG11476 Novel human diagno
5	311	3.2	105	22	ABG12655 Novel human diagno
6	309	3.2	93	22	ABBI0980 Human ovarian and/
7	309	3.2	93	22	AAM96679 Human reproductive
8	308	3.2	107	22	AAU31902 Novel human secret
9	307	3.2	110	23	ABP41552 Human ovarian anti
10	306	3.2	175	22	AAU31725 Novel human secret
11	306	3.2	241	22	ABB11464 Human neuronal thr
12	304	3.1	76	22	AAE75337 Human secreted pro
13	304	3.2	397	17	AAE95913 Neural thread prot
14	300.5	3.1	211	22	ABG12796 Novel human diagno
15	298	3.1	112	21	AAU31789 Human secreted pro
16	294.5	3.0	78	22	ABB11677 Human secreted pro
17	294	3.0	73	22	AAO08565 Human polypeptide
18	294	3.1	304	22	AAU30391 Novel human secret
19	293	3.0	341	22	AAO12925 Human polypeptide
20	293	3.0	119	22	AAE83429 Human immune/haema
21	290	3.0	75	21	AB42504 Human ORFX ORF2268
22	289	3.0	117	23	AAU11247 Human neuronal thr
23	288.5	3.0	136	23	AAE14727 Human Alzheimer's
24	288	3.0	84	22	AAE85552 Human immune/haema
25	288	3.0	108	21	AAU00352 Human secreted pro
26	287.5	3.0	110	21	AAB43292 Human ORFX ORF3056
27	287	3.0	126	23	ABG65578 Human breast speci
28	287	3.0	132	22	ABG08282 Novel human diagno
29	287	3.0	225	22	AAU31713 Novel human secret
30	286	3.0	89	21	AAB43207 Human ORFX ORF2971
31	286	3.0	94	23	ABJ03693 Human ovary specif
32	286	3.0	132	22	AAU69854 Human prostate cDN
33	286	3.0	132	22	AAU01209 Human prostate-spe
34	286	3.0	132	23	ABE95314 Human p776p splice
35	286	3.0	135	22	AAU69829 Human prostate cDN
36	286	3.0	135	22	AAU01284 P789p splice varia
37	286	3.0	135	23	ABE95389 Human P776p splice
38	285.5	3.0	153	22	AAU31783 Novel human secret
39	285	3.0	136	22	AAU78828 Human zinc finger
40	284	2.9	81	22	AAU32132 Novel human secret
41	284	2.9	209	22	AAO12667 Human polypeptide
42	283	2.9	168	22	ABG07628 Novel human diagno
43	282.5	2.9	84	23	ABJ03718 Human ovary specif
44	282.5	2.9	97	23	AAE14726 Human Alzheimer's
45	282	2.9	160	22	AAU78758 Human ribosomal pr

ALIGNMENTS

RESULT 1
ABG68738
ID ABG68738 standard; Protein; 361 AA.
XX ABG68738;
AC ABG68738;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human prostate specific protein DEX0293_107.
XX
KW Human; prostate specific nucleic acid; PSNA; prostate cancer; PSP;
KW prostate specific protein; cytostatic; non-cancerous prostate disease;
KW gene therapy; cancer; immunostimulant; vaccine.
XX
OS Homo sapiens.
XX
PN WO200255735-A2.
XX
PD 18-JUL-2002.

XX PF 27-NOV-2001; 2001WO-US44363.
 XX PR 27-NOV-2000; 2000US-253176P.
 XX PA (DIAD-) DIADEXUS INC.
 XX PI Salceda S, Macina RA, Recipon H, Cafferkey R, Ali S, Sun Y;
 XX PI Liu C, Chen S;
 XX DR WPI: 2002-557831/59.
 XX DR N-PSDB; ABK97631.
 XX PT New prostate specific genes, useful for treating or diagnosing cancer,
 XX PT or useful as vaccines for treating cancer, particularly prostate
 XX PT cancer, in a patient
 XX PS Claim 11; Page 207-208; 212pp; English.
 XX CC The invention relates to a new isolated prostate-specific nucleic acid
 XX CC (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
 XX CC ABK97642 which encode prostate specific proteins appearing as
 XX CC ABG68701-ABG68746, or a sequence hybridising to a PSNA or which has 60%
 XX CC sequence homology with a PSNA. Also included are a method of determining
 XX CC the presence of a PSNA in a sample, a vector comprising the PSNA,
 XX CC a host cell comprising the vector, producing the polypeptide encoded by
 XX CC the PSNA, a method of determining the presence of a PSP in a sample,
 XX CC diagnosing and monitoring the presence and metastases of prostate cancer
 XX CC in a patient, a kit for detecting a risk of cancer or presence of cancer
 XX CC in a patient (the kit comprising a means for determining the presence of
 XX CC the PSNA or PSP in a sample of a patient) and a vaccine comprising the
 XX CC polypeptide or the nucleic acid encoding the polypeptide. The PSNA, PSP
 XX CC and anti-PSP antibody are useful for diagnosing and treating cancer in a
 XX CC patient (e.g. by gene therapy). The nucleic acid molecule and polypeptide
 XX CC are also useful as vaccines for treating cancer, particularly prostate
 XX CC cancer and non-cancerous prostate diseases. The present sequence is
 XX CC a PSP of the invention.
 XX PI Sequence 361 AA;
 XX SQ
 Alignment Scores:
 Pred. No.: 1-54e-30 Length: 361
 Score: 345.50 Matches: 64
 Percent Similarity: 71.8% Conservative: 5
 Best Local Similarity: 66.67% Mismatches: 24
 Query Match: 3.58% Indels: 3
 DB: 23 Gaps: 1
 US-09-966-880A-9 (1-5514) x ABG68738 (1-361)
 QY 5235 TTCAGTTGACATGATCAAGTACTATCTC-----TTTCTTTTTTTTTTTTTTTT 5285
 Db 266 PheTysPheAlaAspSerHisIleGlyLeuAlaPheHisPheAlaPhePhePhePhe 285
 QY 5286 TTTTITTTGAGATGAGTTTGGTCTTCTGCCCATGCTGGATGATGATGATGATGAT 5345
 Db 286 PhePhePheAlaValAlaSerHisProIleAlaGlnAlaGlyValGlnTrpArgAspLeu 305
 QY 5346 AGCTCACTGCACCTCCACCTCCTGGTTTCAAGCAAGCTGTCGCTCAGCTCCCGGT 5405
 Db 306 GlySerLeuGlnProProProGlyPheLysGlnPheLeuCysLeuSerLeuProly 325
 QY 5406 AGATGGGATTACAGCGGCCACCCACCATCTGGCTAATGTTTGTATTTAGTAGAT 5465
 Db 326 SerTrpAspTyrArgAlaProProArgGlnAlaAsnPheCysIlePheSerArgAsp 345
 QY 5466 GGGGTTTCCACCATGTGCGCAGGCTGGTCTCAACTCTGACCTCAGA 5513
 Db 346 GlyValSerProCysTrpTrpGlyTrpSerGlnThrProAspLeuArg 361
 RESULT 2
 AAB43292
 ID AAB43292 standard; Protein; 110 AA.

XX AC AAB43292;
 XX DT 08-FEB-2001 (first entry)
 XX DE Human ORFX ORF3056 polypeptide sequence SEQ ID NO:6112.
 XX KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX OS Homo sapiens.
 XX PN WO2000058473-A2.
 XX XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 XX 02-APR-1999; 99US-0127636.
 XX 05-APR-1999; 99US-0127728.
 XX 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX Shlnkets RA, Leach M;
 DR WPI: 2000-602362/57.
 DR N-PSDB; AAC77501.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 XX useful for treating e.g. cancers, proliferative disorders,
 XX neurodegenerative disorders and cardiovascular disease -
 XX Claim 11; Page 5297-5298; 5507pp; English.
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
 XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
 XX antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 XX antidiabetic; hypotensive; dermatological; immunosuppressive;
 XX antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 XX antithyroid; and antianemic. The sequences can be used for determining
 XX the presence of or predisposition to, or preventing or treating
 XX pathological conditions associated with an ORFX-associated disorder. The
 XX nucleic acids can be used to express ORFX proteins in gene therapy
 XX vectors. The proteins and nucleic acids may be used to treat cancers,
 XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
 XX graft vs host disease, cardiovascular disease, diabetes mellitus,
 XX erythematosus, hypothyroidism, cholesterol ester storage, systemic lupus
 XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 XX allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
 XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
 XX coagulation; to inhibit thrombosis; and as a contraceptive.
 XX Sequence 110 AA;
 Alignment Scores:

```

Pred. No.: 1.84e-29 Length: 110
Score: 334.00 Matches: 62
Percent Similarity: 76.54% Conservative: 0
Best Local Similarity: 76.54% Mismatches: 17
Query Match: 3.46% Indels: 2
DB: 21 Gaps: 2

US-09-966-880A-9 (1-5514) x AAB43292 (1-110)
QY 5274 TTTTTCAGATGGAGTTTGGTCTGTGCCATGCTGGAGTGAA 5333
Db 4 PheCysPheValLeuPheLeuArgTrpSerPheProLeuValAlaGlnAlaGlyVal*** 23
QY 5334 TGGCATGAYCATAGTCACTCAACCTCCACTCTGGGTCAAGCAA---AGCTGTGCGC 5390
Db 24 TrpHisSerLeuGlySerLeuGlnProProLeuProGlyPheLysGlnPheSerCysArg 43
QY 5391 CTCAGCTCCCGGTAGATGGATTACAGGCGCCACCACACACTCGGCTTAATGTTGT 5450
Db 44 ---SerLeuProSerSerTrpAspTyrArgHisAlaProProArgGlnAlaAsnPheCys 62
QY 5451 ATTTTTCAGATGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCCGACCTC 5510
Db 63 IlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrProAspLeu 82
QY 5511 AGA 5513
Db 83 Arg 83

RESULT 3
AAE11962
ID AAE11962 standard; Protein; 202 AA.
AC AAE11962;
XX
XX 18-DEC-2001 (first entry)
DE Human c-myb-related protein #1.
XX
KW Human; c-myb-related gene; thalassaemia; anaemia; haemolytic disorder;
KW haemoglobinopathy; antisense therapy; haematopoietic system disease;
XX aplastic anaemia.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 4
FT /label= Unknown
FT /note= "Encoded by CMC of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 6
FT /label= Unknown
FT /note= "Encoded by TGN of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 9
FT /label= Unknown
FT /note= "Encoded by TAA of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 10
FT /label= Unknown
FT /note= "Encoded by CCN of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 11
FT /label= Unknown
FT /note= "Encoded by TCN of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 19
FT /label= Unknown
FT /note= "Encoded by TAA of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 20
FT /label= Unknown
FT /note= "Encoded by AGN of the inverse complementary

```

```

FT Misc-difference 21 strand of the sequence shown in AAD19493"
FT /label= Unknown
FT /note= "Encoded by CNC of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 32
FT /label= Unknown
FT /note= "Encoded by TAG of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 54
FT /label= Unknown
FT /note= "Encoded by ANT of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 55
FT /label= Unknown
FT /note= "Encoded by TAG of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 62
FT /label= Unknown
FT /note= "Encoded by TAA of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 118
FT /label= Unknown
FT /note= "Encoded by TGA of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 191
FT /label= Unknown
FT /note= "Encoded by TAG of the inverse complementary
FT strand of the sequence shown in AAD19493"
FT Misc-difference 202
FT /label= Unknown
FT /note= "Encoded by TAG of the inverse complementary
FT strand of the sequence shown in AAD19493"
XX
XX WO200172843-A2.
XX
XX 04-OCT-2001.
XX
XX 26-MAR-2001; 2001WO-EP03394.
XX
XX 27-MAR-2000; 2000US-192367P.
XX
XX (FARB ) BAYER AG.
XX
XX Liou J;
XX
XX WPI; 2001-616475/71.
XX N-PSDB; AAD19493.
XX
XX New isolated polynucleotide encoding human c-myb-related polypeptide,
XX for identifying reagents which modulate activity of c-myb-related
XX polypeptide in hematopoietic diseases such as anemia and aplastic
XX anemia -
XX
XX Claim 1; Fig 2; 64pp; English.
XX
XX The present sequence is human c-myb-related protein. The c-myb gene is
XX a proto-oncogene which codes for a nuclear protein expressed primarily
XX in haematopoietic cells. c-myb-related proteins are useful for screening
XX agents which regulate their activity. Human c-myb-related genes can be
XX regulated to prevent or treat diseases of the haematopoietic system
XX such as anaemias including megaloblastic, normochromic and normocytic
XX anaemia, aplastic anaemia, haemolytic disorder, thalassaemia and
XX haemoglobinopathies. Reagents that modulate the activity of c-myb-
XX related protein is useful for treating c-myb-related protein associated
XX disorders. Human c-myb-related gene is also useful in diagnostic assays
XX for detecting diseases and abnormalities related to presence of
XX mutations in the nucleic acid sequences which encode a c-myb-related
XX polypeptide. They are also useful in antisense therapy.
XX
XX Sequence 202 AA;
SQ

```

Alignment Scores:

Pred. No.: 4.32e-29 Length: 202
Score: 332.00 Matches: 73
Percent Similarity: 75.70% Conservative: 8
Best Local Similarity: 68.22% Mismatches: 21
Query Match: 3.45% Indels: 6
DB: 22 Gaps: 3

US-09-966-880A-9 (1-5514) x AAEL1962 (1-202)

QY 3866 TTTCATTATTTTAAATTTTCTTATTTTGGGCAAGGCTGGTCTCT---GNC 3810
DB 78 PheTyrPheLeuPhe---PheLeuLeuPhePhePheTyrGlyArgSerValProLeuVal 96
QY 3809 ACCAGCTGGAGTGCAGTGGCGCAAGCACTGCTCAGTCAGCTTGACCTCTCTGGGCTC 3750
DB 97 AlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProSerProProGlyPhe 116
QY 3749 CAAAGATTCTCCGCTCAGCCC-CCTAGTAGTGGGACACAGGACATGCCACCATGC 3691
DB 117 Lys***PheSerCysLeuSerLeuProSerSerTyrArgHisAlaProCys 136
QY 3690 CCGGCTAATTTTGT---ATAGAAACAGGTTTTCCTATTTGCTCAGGCTGGTCT 3637
DB 137 ProAlaAsn-PheValPheLeuValGluThrGlyPheHisHisIleGlyGlnAlaGlyLe 156
QY 3636 TGAATCCAGGGCTCAAGCATCCACCCACCTCAGCTCCAGTGCCTCCAGTGGATTACAGG 3577
DB 156 uGluLeuGlnThrSerGlyAspProProThrSerAlaSerGlnThrAlaGlyIleThrG 176
QY 3576 CGTGAGCCACACACCTGG 3558
DB 176 yValSerHisArgAlaTrp 182

RESULT 4

ID ABG11476 standard; Protein; 105 AA.
XX AC ABG11476;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #11467.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR N-PSDB; AAS75663.
XX DR WPI; 2001-639362/73.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 41835; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC the polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 105 AA;

Alignment Scores:
Pred. No.: 8.84e-27 Length: 105
Score: 311.00 Matches: 63
Percent Similarity: 68.81% Conservative: 12
Best Local Similarity: 57.80% Mismatches: 26
Query Match: 3.22% Indels: 8
DB: 22 Gaps: 3

US-09-966-880A-9 (1-5514) x ABG11476 (1-105)

QY 5199 TTTTAAATTTAGAGCTCTTCTATTTTCCATAGTATTCAAGTTTCACATCATCAAGTA 5258
DB 1 PheLeuSerLeuProThrPheLeuPhe---ValIlePheSerGlyGluGluLeuLeu 19
QY 5259 TTACTCTTTCTT-----TTTTTTTTTTTTTTTTTTTTTTTTCAGATGGAGTTT 5306
DB 20 ValLeuAlaLeuValPheLeuSerLeuPhePhePhePheLeuArgTrpSerPhe 39
QY 5307 GGTCTTTGTCCTGAGTGGAGTGGATGGATGATGATGATGATGATGATGATGATGATG 5366
DB 40 AlaValAlaAlaGlnAlaValValGlnTrpHisAsnLeuSerSerLeu-----Pro 56
QY 5367 CCGTGGTTCAAGCAAGCTGTGCGCTCAGCTCCCGCTAGATGGATGATGATGATGATG 5426
DB 57 PheGlyPheLysGlnPheSerCysLeuSerLeuProSerSerTrpAspTrpArgCysPro 76
QY 5427 CCACCACATCGGCTAATTTTGTATTTTATAGTAGATGGGTTTTCACCATGTTGGCCA 5486
DB 77 ProProArgProAlaAsnPheCysIlePheThrArgAspGlyValSerProCysCysPro 96
QY 5487 GCGTGGTCTCAACTCTCTGACCTCAGA 5513
DB 97 GlyTrpSerArgThrSerAspLeuArg 105

RESULT 5

ID ABG12655
XX ABG12655 standard; Protein; 105 AA.
XX AC ABG12655;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #12646.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX

PR	06-SEP-2000;	2000US-02304333
PR	08-SEP-2000;	2000US-02312422
PR	08-SEP-2000;	2000US-02312431
PR	08-SEP-2000;	2000US-02312444
PR	08-SEP-2000;	2000US-02313413
PR	08-SEP-2000;	2000US-02314114
PR	08-SEP-2000;	2000US-02320800
PR	08-SEP-2000;	2000US-02320811
PR	12-SEP-2000;	2000US-02319688
PR	14-SEP-2000;	2000US-02323977
PR	14-SEP-2000;	2000US-02323989
PR	14-SEP-2000;	2000US-02323998
PR	14-SEP-2000;	2000US-02324000
PR	14-SEP-2000;	2000US-02324011
PR	14-SEP-2000;	2000US-02330633
PR	14-SEP-2000;	2000US-02330644
PR	21-SEP-2000;	2000US-02330655
PR	21-SEP-2000;	2000US-02344223
PR	21-SEP-2000;	2000US-02344277
PR	25-SEP-2000;	2000US-02349977
PR	25-SEP-2000;	2000US-02349988
PR	26-SEP-2000;	2000US-02354884
PR	27-SEP-2000;	2000US-02358834
PR	29-SEP-2000;	2000US-02358836
PR	29-SEP-2000;	2000US-02363627
PR	29-SEP-2000;	2000US-02363677
PR	29-SEP-2000;	2000US-02363688
PR	29-SEP-2000;	2000US-02363699
PR	29-SEP-2000;	2000US-02363770
PR	02-OCT-2000;	2000US-02368002
PR	02-OCT-2000;	2000US-02370737
PR	02-OCT-2000;	2000US-02370738
PR	02-OCT-2000;	2000US-02370739
PR	13-OCT-2000;	2000US-02370404
PR	13-OCT-2000;	2000US-02399935
PR	13-OCT-2000;	2000US-02399937
PR	20-OCT-2000;	2000US-02409660
PR	20-OCT-2000;	2000US-02412211
PR	20-OCT-2000;	2000US-02417885
PR	20-OCT-2000;	2000US-02417886
PR	20-OCT-2000;	2000US-02417887
PR	20-OCT-2000;	2000US-02418008
PR	20-OCT-2000;	2000US-02418009
PR	20-OCT-2000;	2000US-02418236
PR	01-NOV-2000;	2000US-02446117
PR	08-NOV-2000;	2000US-02464174
PR	08-NOV-2000;	2000US-02464675
PR	08-NOV-2000;	2000US-02464676
PR	08-NOV-2000;	2000US-02464777
PR	08-NOV-2000;	2000US-02464778
PR	08-NOV-2000;	2000US-02465233
PR	08-NOV-2000;	2000US-02465234
PR	08-NOV-2000;	2000US-02465235
PR	08-NOV-2000;	2000US-02465236
PR	08-NOV-2000;	2000US-02465237
PR	08-NOV-2000;	2000US-02465238
PR	08-NOV-2000;	2000US-02465332
PR	08-NOV-2000;	2000US-02466609
PR	08-NOV-2000;	2000US-02466610
PR	08-NOV-2000;	2000US-02466611
PR	17-NOV-2000;	2000US-02492111
PR	17-NOV-2000;	2000US-02492112
PR	17-NOV-2000;	2000US-02492113
PR	17-NOV-2000;	2000US-02492114
PR	17-NOV-2000;	2000US-02492115
PR	17-NOV-2000;	2000US-02492116
PR	17-NOV-2000;	2000US-02492117
PR	17-NOV-2000;	2000US-02492118

PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249246.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251198.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251749.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-)	HUMAN GENOME SCI INC

Rosen CA, Barash SC, Ruben SM:

WPI; 2001-488786/53.

WF-I, ZOOI-488786/
N-PSDB: ABA07691.

New isolated ovarian and/or breast cancer related nucleic acids and polypeptides, useful for diagnosing, treating and/or preventing human diseases and disorders, particularly ovarian and/or breast cancer -

Claim 11; SEQ ID NO 486; 577pp + Sequence Listing; English.

The invention relates to novel genes (ABA07454-ABA08224) and proteins (ABB10743-ABB10980) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 93 AA:

Alignment Scores:

Assignment Scores:

Pred. No.:	1.42e-26
Score:	309.00
Percent Similarity:	76.25%
Best Local Similarity:	71.25%
Query Match:	3.20%
DB:	22

US-09-966-880A-9 (1-5514) x ABB10980 (1-93)

[illegible]

QY	5333	ATGGCATGAYCATAGTCTACTGCAACCTCCACCTCTCTGGGTTCAGCAAGAGCTGTGCCT	5392
Db	23	ntrpHisTyrLeuSerSerLeuGlnProProAargPheLysGlnPheSerCysLe	43

QY	5393	CAGCCTCCGGGTAGATGGGATTACAGCGGCCACCACACACTCGGCTAATGTTGTAT	5452	PR	08-SEP-2000;	2000US-0231242.
Db	43	userLeuSerSerTyrArgHisAlaAlaAsnPheCys11	63	PR	08-SEP-2000;	2000US-0231243.
QY	5453	TTTTAGTAGATGGGTTTCACCATGTTGGCCAGGCTGGTCAAACTCCTGACCTC	5510	PR	08-SEP-2000;	2000US-0231244.
Db	63	epheserArgAspGlyValSerProCysTrpSerGlyTrpSerArgThrProAspLeu	82	PR	08-SEP-2000;	2000US-0231413.
RESULT 7				PR	08-SEP-2000;	2000US-0231414.
AA96679				PR	08-SEP-2000;	2000US-0232080.
ID	AA96679	standard; Protein; 93 AA.		PR	12-SEP-2000;	2000US-0232081.
AC	AA96679;			PR	14-SEP-2000;	2000US-0232397.
XX				PR	14-SEP-2000;	2000US-0232398.
DT	21-NOV-2001	(first entry)		PR	14-SEP-2000;	2000US-0232399.
XX				PR	14-SEP-2000;	2000US-0232400.
DE		Human reproductive system related antigen SEQ ID NO: 5337.		PR	14-SEP-2000;	2000US-0232401.
XX				PR	14-SEP-2000;	2000US-0233063.
KW		Human; reproductive system related antigen; reproductive system disorder;		PR	14-SEP-2000;	2000US-0233064.
XX		cancer; gene therapy.		PR	14-SEP-2000;	2000US-0233065.
OS		Homo sapiens.		PR	14-SEP-2000;	2000US-0233066.
XX				PR	21-SEP-2000;	2000US-0234223.
PN	WO200155320-A2.			PR	21-SEP-2000;	2000US-0234274.
XX				PR	25-SEP-2000;	2000US-0234997.
PD	02-AUG-2001.			PR	25-SEP-2000;	2000US-0234998.
XX				PR	26-SEP-2000;	2000US-0235484.
PF	17-JAN-2001;	2001WO-US01339.		PR	27-SEP-2000;	2000US-0235834.
XX				PR	27-SEP-2000;	2000US-0235836.
PR	31-JAN-2000;	2000US-0179065.		PR	29-SEP-2000;	2000US-0236327.
PR	04-FEB-2000;	2000US-0180628.		PR	29-SEP-2000;	2000US-0236367.
PR	24-FEB-2000;	2000US-0184664.		PR	29-SEP-2000;	2000US-0236368.
PR	02-MAR-2000;	2000US-0186350.		PR	29-SEP-2000;	2000US-0236369.
PR	16-MAR-2000;	2000US-0189874.		PR	29-SEP-2000;	2000US-0236370.
PR	17-MAR-2000;	2000US-0190076.		PR	02-OCT-2000;	2000US-0236802.
PR	18-APR-2000;	2000US-0198123.		PR	02-OCT-2000;	2000US-0237037.
PR	19-MAY-2000;	2000US-0205515.		PR	02-OCT-2000;	2000US-0237038.
PR	07-JUN-2000;	2000US-0209467.		PR	02-OCT-2000;	2000US-0237039.
PR	28-JUN-2000;	2000US-0214886.		PR	02-OCT-2000;	2000US-0237040.
PR	30-JUN-2000;	2000US-0215135.		PR	13-OCT-2000;	2000US-0239935.
PR	07-JUL-2000;	2000US-0216647.		PR	13-OCT-2000;	2000US-0239937.
PR	07-JUL-2000;	2000US-0216880.		PR	20-OCT-2000;	2000US-0240960.
PR	11-JUL-2000;	2000US-0217487.		PR	20-OCT-2000;	2000US-0241221.
PR	11-JUL-2000;	2000US-0217496.		PR	20-OCT-2000;	2000US-0241785.
PR	14-JUL-2000;	2000US-0218290.		PR	20-OCT-2000;	2000US-0244617.
PR	26-JUL-2000;	2000US-0220963.		PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000;	2000US-0220964.		PR	08-NOV-2000;	2000US-0246475.
PR	14-AUG-2000;	2000US-0224518.		PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224519.		PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0225213.		PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225214.		PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225266.		PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225267.		PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225268.		PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225270.		PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225447.		PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225757.		PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225758.		PR	08-NOV-2000;	2000US-0246609.
PR	18-AUG-2000;	2000US-0225759.		PR	08-NOV-2000;	2000US-0246610.
PR	22-AUG-2000;	2000US-0226279.		PR	08-NOV-2000;	2000US-0246611.
PR	22-AUG-2000;	2000US-0226681.		PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226868.		PR	17-NOV-2000;	2000US-0249207.
PR	23-AUG-2000;	2000US-0227182.		PR	17-NOV-2000;	2000US-0249208.
PR	30-AUG-2000;	2000US-0227009.		PR	17-NOV-2000;	2000US-0249209.
PR	01-SEP-2000;	2000US-0228924.		PR	17-NOV-2000;	2000US-0249210.
PR	01-SEP-2000;	2000US-0229287.		PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229343.		PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229344.		PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229345.		PR	17-NOV-2000;	2000US-0249214.
PR	05-SEP-2000;	2000US-0229509.		PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229513.		PR	17-NOV-2000;	2000US-0249216.
PR	06-SEP-2000;	2000US-0230437.		PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230438.		PR	17-NOV-2000;	2000US-0249218.
				PR	17-NOV-2000;	2000US-0249244.

```
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
DR N-PSDB; AAL02649.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX
PS Claim 11; SEQ ID NO 5337; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.
XX
SQ Sequence 93 AA;
Alignment Scores:
Pred. No.: 1,42e-26 Length: 93
Score: 309.00 Matches: 57
Percent Similarity: 76.23% Conservative: 4
Best Local Similarity: 71.23% Mismatches: 18
Query Match: 3.20% Indels: 1
DB: 22 Gaps: 0
US-09-966-880A-9 (1-5514) x AAM96679 (1-93)
QY 5273 TTTTNTTTTTTTTTTTTTCAGATGGAGTTTGTGCTTGTGCTCCATGCTGCCATGCTGGAGTGA 5332
Db 4 PhePhePhePhePhePhePheGluThrGluSerHisSer-ValThrGlnAlaGlyIleGl 23
QY 5333 ATGGCATGAYCATAGTCACTCACTCAACCTCCACCTCCTGGTTCAAGCAAGCTGCGCT 5392
Db 23 ntrpHisTyLeuSerSerLeuGlnProProProProArgPheLysGlnPheSerCysLe 43
QY 5393 CAGCCTCCGGGTAGATGGATTACAGCGCCACACACACACTCGGCTAAATGTTGTAT 5452
Db 43 uSerLeuLeuSerSerTrpAspTyArgHisThrProProHisAlaAlaAsnPhcCysIl 63
QY 5453 TTTTATGATAGATGGGTTTACCATGTTGGCCAGCGTGGTCTCAAACTCCCTGACCTC 5510
Db 63 ePheSerArgaspGlyValSerProCysTrpSerGlyTrpSerArgThrProAspLeu 82
RESULT 8
AU31902
ID AU31902 standard; Protein; 107 AA.
XX
AC AU31902;
XX
XX 18-DEC-2001 (first entry)
XX
```

```
DE Novel human secreted protein #2393.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 524; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 107 AA;
Alignment Scores:
Pred. No.: 2e-26 Length: 107
Score: 308.00 Matches: 60
Percent Similarity: 77.50% Conservative: 2
Best Local Similarity: 75.00% Mismatches: 17
Query Match: 3.19% Indels: 1
DB: 22 Gaps: 0
US-09-966-880A-9 (1-5514) x AAU31902 (1-107)
QY 5274 TTTTNTTTTTTTTTTTTTCAGATGGAGTTTGTGCTTGTGCTCCATGCTGCCATGCTGGAGTGA 5333
Db 3 PhePhePhePhePhePhePheGluThrLeuValAlaGlnAlaGlyVal*** 22
QY 5334 TGGCATGAYCATAGTCACTCACTCAACCTCCACCTCCTGGTTCAAGCAAGCTGTCGCCTC 5393
Db 23 LeuHisAspProGlu***LeuGlnProProProProGlyPheLysArgPheSerCysLeu 42
QY 5394 AGCCTCCCGGTAGATGGATTACAGCGCCACACACACTCGGCTAAATGTTGTAT 5453
Db 43 SerLeuProSerSerTrpAspTyArgHisProProProProProAlaAsn-Phc***Ph 62
QY 5454 TTTAGTAGATGGGTTTTCACCATGTTGGCCAGCGTGGTCTCAAACTCCCTGACCTCA 5511
Db 62 eLeuValGluThrGlyPheHisValGlyGlnAlaGlyLeuGluLeuLeuThrSer 81
```

RESULT 9
ABP41552
ID ABP41552 standard; Protein: 110 AA.
XX
AC ABP41552;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HSPSI60, SEQ ID NO:2684.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antinflammatory; gynaecological; reproductive.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.
XX
DR N-PSDB; ABQ54629.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 11; SEQ ID NO 2684; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis), systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 110 AA;
XX
Alignment Scores:
Pred. No.: 2.66e-26 Length: 110
Score: 307.00 Matches: 72
Percent Similarity: 72.48% Conservative: 7
Best Local Similarity: 66.06% Mismatches: 22
Query Match: 3.19% Indels: 9
DB: 23 Gaps: 3
XX
US-09-966-880A-9 (1-5514) x ABP41552 (1-110)
XX
QY 3865 TTCCATTTATTTTAAATTTT-----CTTCTTATTTTGGAGCAAGTCTG 3818
DB 3 PhePhePhePhePheLeuPheSerAspValIleLeuTyrLeuPheLeuArgGlnAsn--- 21
QY 3817 GTCTGTGTCACCCAGGCTGGAGTGCAGTGGCGCAAGCACTGCTCAGTGCAGCCTTGACCTC 3758
DB 22 ---LeuSerProArgLeuGluCysSerGlyMetIleSerAlaHisCysArgLeuTyrLeu 40
QY 3757 CTGGGCTCCAAAGATTCTCTGCTGCTACGCCCTTA-GTAGCTGGGACGACAGGCATGC 3699
DB 41 MetGlySerSerAspSerProAlaSerAlaSerGlnValValGlySerThrGlyThrHis 60
QY 3698 CACCATGCCCGCTAAATTTTGT-----ATAGAAACAGGGTGTTCATCTGCTCAG 3645
DB 61 HisHisThrGlnLeuIlePheValPheLeuValGluMetGlyPheHisValAspGln 80
QY 3644 GCTGCTTGAACCTCCAGGCTCAAGCATCCACCCACTCAGCCTCCAGAGTGTCTGG 3585
DB 81 AlaGlyLeuGluLeuLeuThrSerSerAspProThrSerAlaSerGlnSerAlaGly 100
QY 3584 ATTACAGCGTGAGCCACACACCTGG 3558
DB 101 IleThrGlyValSerHisCysThrTrp 109
XX
RESULT 10
AAU31725
ID AAU31725 standard; Protein: 175 AA.
XX
AC AAU31725;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2216.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
XX
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX

polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

AA	Sequence	211 AA;
SQ		

Alignment Scores:	2.15e-25	Length:	211
Pred. No.:	300.50	Matches:	57
Score:	63.27%	Conservative:	33
Percent Similarity:	58.16%	Mismatches:	33
Best Local Similarity:	3.11%	Indels:	1
Query Match:	22	Gaps:	1
DB:			

US-09-966-880A-9 (1-5514) x ABG12796 (1-211)

Qy	5220	TATTTTCCATAGATTCAAGTTTGACATGATCAAGTATTACTCTTTCTCTTTCTTTT	5279
Db	112	YrrLeuProtyrIleGlySerCysGlnGlyProSerLeuPheSerPheSerPhePhe	131
Qy	5280	TTTTTTTTTTTTTGTG-----AGATGAGAGTTTGGTGTGTTGGCCCATGCTGGAGTG	5330
Db	132	PhePhePhePheLeuArgArgArgThrGluSerCysSerValAlaGlnValGlyVal	151
Qy	5331	GAATGGCATGAYCATAGCTCACTGCACACTCCACTCTCTGGTTTCAGCAAGCTGTGCG	5390
Db	152	GlnTrpArgAspLeuGlySerLeuGlnProProProGlyPheLysArgPheSerCys	171
Qy	5391	CTCAGCCTCCCGGGTAGATGGGATTACAGCGCGCCACACACACTCGGCTAATGTTGT	5450
Db	172	LeuSerLeuProSerSerTrpAspTyrArgArgAlaProProArgProAlaAsnPheTyr	191
Qy	5451	ATTTTDTAGTAGATGGGGTTTCAACATGTGGCCAGGTGGTCTCAAACTCTCT	5504
Db	192	IlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerArgTrpPro	209

RESULT 15

RESULTS
AAG03789
ID AAG03789 standard; Protein: 112 AA.

DT 06-OCT-2000 (first entry)

xx Human secreted protein, SEQ ID NO: 7870.

AA Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.

OS Homo sapiens.

PN EP1033401-A2.

XX
PD
06-SEP-2000.

XX
PF
21-FEB-2000: 2000EP-0200610.

XX	26-FEB-1999;	99US-0122487.
PR	(GEST)	GENSET.
XX		
PA	Dumas Milne Edwards J, Duclert A, Giordano J;	
XX		
PI	WPI; 2000-500381/45.	
DR	N-PSDB; AAC03795.	
XX		
XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -	
XX		
PS	Claim 13; SEQ ID 7870; 71pp + CD-ROM; English.	
XX		
CC	The present sequence is a polypeptide encoded by one of a large number	
CC	of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs	
CC	were prepared from total human RNAs or polyA+ RNAs derived from 30	
CC	different tissues. EST sequences usually correspond mainly to the 3'	
CC	untranslated region (UTR) of the mRNA because they are often obtained	
CC	from oligo-dT primed cDNA libraries. Such ESTs are not well suited for	
CC	isolating cDNA sequences derived from the 5' ends of mRNAs and even in	
CC	those cases where longer cDNA sequences have been obtained, the full 5'	
CC	UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'	
CC	ends and can therefore be used to obtain full length cDNAs and genomic	
CC	DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and	
CC	chromosome mapping procedures. They are used to obtain upstream	
CC	regulatory sequences and to design expression and secretion vectors.	
XX		
SQ	Sequence	112 AA;

Alignment Scores:		
Pred. No.:	3.04e-25	Length:
Score:	298.00	Matches:
Percent Similarity:	73.08%	Conservative:
Best Local Similarity:	66.35%	Mismatches:
Query Match:	3.10%	Indels:
DB:	21	Gaps:

US-09-966-880A-9 (1-5514) x AAG03789 (1-112)

3853	Qy	TTAATTTTCTCTTATTTTTTTGAGCAAGGTCTGGTCTGTACCCAGGCTGGAGTGC	3794
		: : : : : : : : : : : :	
8	Db	ValValValValValValLeuArgGlnSerLeuAlaLeuLeuProArgLeuGluCys	27
3793	Qy	AGTGGCGCAGCACTGCTCAGTGCAGCCTTGACCTCTCTGGGTCCAAAGATTCTCTGCC	3734
		: : : : : : : : : : : :	
28	Db	SerGlyThrIleLeuAlaHisCysSerLeuCysLeuGlySerSerAspSerProAla	47
3733	Qy	TCAGCCCCCTA-CTAGTGGACCACAGGCACATGCCACCTCCCGCGCTAATTTTGGT	3675
		: : : : : : : : : : : :	
48	Db	SerAlaSerGlnValAlaArgThrThrGlyMetCysHisHisThrGlnLeuIlePheVal	67
3674	Qy	-----ATAGAAACAGGGTTTGGCCATGTGGCTCAGGCTGGTCTTGAATCCAGGCTCA	3621
		: : : : : : : : : : : :	
68	Db	PheLeuIleGluThrGlyPheLeuHisIleGlyGlnAlaLeuGluLeuLeuThrSer	87
3620	Qy	AGCGATCCACCCACCTCAGCCTCCCAAGATGGCTGGGATTACAGGCGGTGAGCCACACACC	3561
		: : : : : : : : : : : :	
88	Db	GlyAspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyVal-***ThrThrPr	107
3560	Qy	TGGCTGTCTG	3551
		: : : : : : : : : : : :	
107	Db	oGlyArgLeu	110

Search completed: June 14, 2003, 18:25:52
Job time : 248.945 secs

A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the

Alignment Scores:

Pred. No.: 9.59e-38 Length: 673
Score: 445.00 Matches: 180
Percent Similarity: 32.32% Conservative: 44
Best Local Similarity: 25.97% Mismatches: 161
Query Match: 4.62% Indels: 309
DB: 4 Gaps: 22

US-09-966-880A-9 (1-5514) x F40201 (1-673)

```

QY 5510 GAGTCAGGAGTTTGACACAGCCTGGCCACATGGTGAACCCCATCTCTACTATAAAT 5451
D 5510 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 248 GluValArgSerLeuArgProAla***SerThrTrp***AsnCysValSerIleLysAsn 267
QY 5450 ACRAACATTAGCCGAGTGTGGTGGCGCCTGTAATCCCATCTACCCGGGAGCTGAG 5391
D 5450 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 268 Thr***IleSerGlnGluTrpArgProMetProValIleProAlaThrArgGluThrGlu 287
QY 5390 GCGACAGCTTTGGTTGAACCCAGGAGGTGGAGGTTCGAGTCAGTCATGCCATTC 5331
D 5390 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 288 AlaglyGluSerLeuGluProArgArgLeuArgLeuGln***Ala-LysThrCysHisCy 307
QY 5330 CACTCCAGCATGGGCACACAGACCAAACTCCATCTCAAAAAAATAAAAAA 5271
D 5330 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 307 sthrProAla-GlyValSerGlnGlyProIleSerLysThrLysGlnLysL 327
QY 5270 AAGAAAGAGTAATCTTGATCATTTGTCAAACCTTGAATACATGGAATAAGAGACTC 5211
D 5270 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 327 yslsLys-----LysAsnProGlnLys- 334
QY 5210 TAAATTTTAAATATTAAGATTTTTAAATCTCAAATATCATAGTCTCTAATAATA 5151
D 5210 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 334 ----- 334
QY 5150 ATTAGACTTAGTGTCTAATGACAAATGACTGTTAAGATTCTTGCTTTTATTCCTCTAGTCT 5091
D 5150 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 334 ----- 334
QY 5090 TTCTACAGCAGCATATTTTACATGATTGG-----ATTATCTGATTGCATGTGATT 5037
D 5090 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 335 -----Gln*****PheCysPheTrpGlyPhePhePhePheCysPheValP 353
QY 5036 TGTACTCTGCTGCTTTTAACTAGTCAAAATGATCTAAATATCTTGTCTGTTGTTCCC 4977
D 5036 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 353 heVal----- 354
QY 4976 ACATAATGACTGATTTCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4917
D 4976 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 354 ----- 354
QY 4916 TCATAAGCACTAATTTGAATAGCATTTGTGGAATACTCTCGTAAACCAAGTTTCCCTAA 4857
D 4916 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 354 ----- 354
QY 4856 AGAAATCAAGAGTCCCATTTGCAGCTGCAGACAGCAACAAGCTGTACAGCGTGGTTGA 4797
D 4856 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 355 -----PheGluIleGlyProCysSerAspThrProAlaGlyValGlnTrpGlnV 371
QY 4796 TATGAAGACACAGTAGTGGCTGGATTGTTGCCCTAGCTTTTCCAGAAATGCAATTATC 4737
D 4796 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 371 allLeuAlaHis-----CysSerLeuA 378
QY 4736 CAATTTGTGGAGCTAGTCTGTACCTAATGGTACGGGGCAGAGTGCAGAGAAATTTGCG 4677
D 4736 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 378 snLeuGlySerSerAspSerPro-----AlaSerValSerArgValAlag 394
QY 4676 AAGTAAAGGGCTCAGGAATTTGAATAACGTTTGTGCGCAAGAGAAATTTTCTTA 4617

```

```

D 394 lyIleThrGlyMetGly----- 399
QY 4616 CACATTATTTTCTTACATTTAGGTATATAATTTACATACAGTAACCTCCACCCATTAAAG 4557
D 399 ----- 399
QY 4556 ATGCATAATTTCTAGCAATATTTTGGATGTAAACACACCCAACTGCTTACCACAGTCAGGA 4497
D 400 -----ArgHisSerTrp----- 403
QY 4496 TACAGAACTGAATCACCACCTCACCCACAGAGCTTCTTCCCTCTTTGTCAGTCCTTC 4437
D 403 ----- 403
QY 4436 CCCTCCCCACAGGACAACCACTGATTTGCTTTCTTTTCCCAATTTTGTGGAGAATGG 4377
D 404 -----LeuIleTyrValPheLeuIleGluThrG 413
QY 4376 AGTTTCACTGTGTGGCCAGGCTGTGAACCTCTGAGTCTAGCAATCTTCCACAGT 4317
D 413 lnPheHisHisValasGlnAlaGlyLeuLysLeuLeuThrSerSerAspLeuProSerT 433
QY 4316 GGCTCTCTAAAGTCTGGGATTTACAGGTGTGAGCTCTATGCTGCTGCTGATTTGCTTTC 4257
D 433 rpSerProLysValLeuGly***GlnAla***AlaThrThrProSer***** 449
QY 4256 TGTAACTATAAGTTACTTTTCATTTTCTAGAAATTTTGTATGGCATACATATCTCTCCTT 4197
D 450 -----PheValPheGlyGlyPhePhePhePhePheAla----- 463
QY 4196 CATGCTGCTGCTCTTTCATTTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 4137
D 464 -----LeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArg----- 477
QY 4136 CTTTGTATGTCAGGCTGATTTCCATTTGCATGATATAGCACATTTTGTATTGTTTCA 4077
D 478 --LeuGluCysSerGlyLysPhe----- 484
QY 4076 CTGTTGATGATATTTGGATAGTCTCTACTTTGGGGTTTGTGTAATAAGGTACTATGA 4017
D 485 -----TrpLeuThrAlaAlaSerThrSerTrp-----ValGlnA 496
QY 4016 ACATTACGTCACAGTCTTT-----GTCTGGATATACATTTGAGA 3978
D 496 laIleLeuLeuProLeuSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaIleProG 516
QY 3977 AT-----TGATATTTT-----A 3966
D 516 ly***PheMetTyrPhe*****ArgHisSerPheThrMetLeuIleArgLeuValSerA 536
QY 3965 ATAATGTAGTCAGGTGACAGACCCCTGACTTTTAAAGTTAAATACCAAAA----- 3916
D 536 snSer***ProGlnValIleCysProGly-----LeuProLysCysTrpAspA 553
QY 3915 ----TATCAGCTTAACCTAAGTACTCATCTAGGACACACAGCTCTTTGTAGTCTGTTTCCA 3861
D 553 spArgArgGluProHisProAla-----*****LeuPheLeuGlyV 569
QY 3860 TTTATTTTAAATTTTCTTCTTATTTTGTAGGCAAGGTCTGCTCTCTG----- 3811
D 569 alPhePhePhePhePheLeuLeuCysPhe-----CysPhe***AspA 583
QY 3810 -----CACCCAGCTGGAGTGCAGTGGCGCAAGCACTGCTAGTCAGCGCCT 3765
D 583 rgProLeuLeu***HisProGlyTrpSerAlaValAla-SerPheGlySerLeuGlnPro 602
QY 3764 TGACCTCTGGCTCCAAAGATCTCTGCTGCTCAGCCC-CCTAGTACTGGGACACAGG 3706
D 603 GlnProGlyPheLysArgPheSerCysLeuCysLeuProCysSerTrpAspTrpArg 622
QY 3705 CACATGTCACCATCCCGGCTAATTTTGTGTTATAGAAACAGG-----GTTTTGCAATGT 3652

```

Db 623 HisGlyProPheLeuAlaAsnLeuCysIlePheAsnArgAspThrValSerProCys 642
QY 3651 TGCTCAGGCTGGTCTTGAACCTCAGGGCTCAAGCGATCCACCCAGCTCCAGCAG 3592
Db 643 ***SerGlyTrpSerGlnThrProAspLeuLys***SerAlaLeuLeu-ValSerGlnSe 662
QY 3591 TGCTGGATTACAGCGGTGAGCCACACACC 3561
Db 662 rAlaGlyMetThrGlyValSerHisHisThr 672
RESULT 3
F40201
artifact-warning sequence (translated ALU class F) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: F40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: F40201
A:Molecule type: DNA
A:Residues: 1-673 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; PMID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th
Alignment Scores:
Pred. No.: 4,7e-36 Length: 673
Score: 429.00 Matches: 204
Percent Similarity: 36.04% Conservative: 71
Best Local Similarity: 26.74% Mismatches: 268
Query Match: 4.44% Indels: 221
DB: 4 Gaps: 24
US-09-966-880A-9 (1-5514) x F40201 (1-673)
QY 3562 GGTGGTGGCTCAGCGCTGTATCCAGCAGCTCTGGAGGCTGAGTGGGTGATCGCTT 3621
Db 3 ValTrpLeuThrProValIleProAlaLeuTrpGluThr-ArgArgAlaAspHisLe 22
QY 3622 GAGCCCTGGAGTCAAGACCCAGCTGAGCAACATGCAAAACCCCTGTTCTATACAA-- 3679
Db 22 uArgSerGlyVal***AspGlnProAspGlnHisGlyGluThrValSerLeuLeuLysIl 42
QY 3680 ----AAATTAGCCGGGCATGGTGGCATGTGCTGGTCCCGCTACTAGGGGG-CTGAG 3734
Db 42 eHisLysLeuAlaArgAsnGlyGlyProCysLeu***SerGlnLeuHisGlyArgGlnAr 62
QY 3735 GCAGGAGATCTTTGGAGCCAGGAGGTCAAGCGTGCAGTGCAGTGTCTGGCCACGTG 3794
Db 62 gGlnGluAsnArgLeuAsnProGlyGly***GlyCysSerGluProLysLeu-AlaThrA 82
QY 3795 CACTCCAGCTGGTGACAGGACAGCTCGCTCCCTCAAAATAAAGAGAAAAATTA 3854
Db 82 lLeuGlnProGlyCysGlnSerLysGlyLeuSerGlnLysGlnLysLysL 102
QY 3855 AATAAATGAAACA-----ACTACAAAGAGCTGTGTCTAGTAGAGCTA 3899
Db 102 yLysLysLysThrProLysAsnLys*****AlaGlyCysGlyGlySerArgLeuS 122
QY 3900 CTTAGTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGCTGTCACCTGC- 3955
Db 122 er-----SerGlnHisPheGly-----ArgProGlyGlyGlnIleThr***GlyG 137
QY 3956 -----ACTACATTATTAATAAT------CAATCTCAATGTATATCC 3992
Db 137 InGluPheCluThrSerLeuIleAsnMetValLysLeuCysLeuTyrr***LysTyrlleA 157

QY 3993 ACACAAAGACTGGTACGTGAATGTTTCATAGTACCTTTATTCACAAACCCCAAGTAGAG 4052
Db 157 sn***ProGlyMetAlaAlaHisAlaCysAsnProSerTyrrThrGlyAspArgGlyArgA 177
QY 4053 ACTATCCCAANTATCCATCAACAAGTGAACAATAAACAANAATGTGTATATCC-----A 4106
Db 177 rgIleAla***ThrGlnGlu-----ValGluAlaAlaValSerGlnAsnL 192
QY 4107 TGCATGGGAATACCCACCCTGCAGTCAAAAGGAAGACTTACTTGGGGATCAATCCCAAG 4166
Db 192 euProLeuHisSerSerArgGlyValArgAlaAlaTyrrLeuLysAsnLysAla 212
QY 4167 TCATGACGCTAAATGAAGAGTGCAGACATCAAGGAGGAGATAATGTATGCCATACCAAT 4226
Db 212 laLys----- 213
QY 4227 TCTAGAAATGAAAGTAACCTTATAGTTACAGAAAGCAAAATCAGGGCAGGATAGAGCTC 4286
Db 214 -----LysLysLysLysLysProLysThrLys*****LeuGlyValAlaLah 232
QY 4287 ACACCTGTAATCCAGCAGCTTTGAGAGGCCAGCTGGGAAGATTGCTAGAACTCAGCAGATT 4346
Db 232 lAlaCysHisProSerThrLeuGlyAspGlnCluGlyArgSerLeuGluValArgSerL 252
QY 4347 CAAGCAGCCTGGGCAACACAGTGAACCTCTCCACAAAAATGGGAAAAAAGAA 4406
Db 252 euArgProAla***SerThrTrp***AsnCysValSerIleLysAsnThr***IleSerG 272
QY 4407 AGCAATCAGTGGTGTCTCTGTG-----GGAGGGGAA 4439
Db 272 InGluTrpArgProMetProValIleProAlaThrArgGluThrGluAlaGlyLysL 292
QY 4440 GGACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGGTGGTATTCAGTCTGTGATCC 4499
Db 292 euGluProArgArgLeuArgLeuGln***Ala-----LysThrCysHisC 307
QY 4500 TGACTGTGTAGCAGTTTGGGGTGTTCACATCCAAAAATATTCTGTAAGAATTATGATCATCT 4559
Db 307 ysThrProAlaGlyVal-----SerG 314
QY 4560 AAATGGTGGAGTTTACTCTATGTAAATATACCTCAATGTAGAAAAAATAATGTGTAA 4619
Db 314 luGlnGlyProIle-SerLysThrLysThrLysGlnLysLysLysLysAsnProGln 333
QY 4620 GAAAGCTTTCAATCTCTTCCAGCAACAGTTATTCAAATTCCTGAGCCCTTTACTTCGC 4679
Db 334 LysGln*****PheCysPheTrpGlyPhePhe----- 345
QY 4680 AAATTCTCTGCATCTCTGCCCGTA-----CCATTAGGTGACAGCAC 4721
Db 346 ---Phe-PhePheCysPheValPheValPheGluIleGlyProCysSerAspThrPr 364
QY 4722 TAGCTCCCAAAATTTG----- 4737
Db 364 oAlaGlyValGlnTrpGlnValLeuAlaHisCysSerLeuAsnLeuLeuGlySerAs 384
QY 4737 ----- 4737
Db 384 pSerProAlaSerValSerArgValAlaGlyIleThrGlyMetGlyArgHisSerTrpLe 404
QY 4738 -ATAAATGCTTTCTGGAAAAGACTAGG-----GACAAAATCCAGGCTACTTGTGTCT 4790
Db 404 uIleTyrrValPheLeuIleGluThrGlnPheHisHisValAspGlnAlaGlyLeuLysLe 424
QY 4791 T-----TCATATCAACACCGCTCTACAGCTTGTGTGTCTGTCTG 4829
Db 424 uLeuThrSerSerAspLeuProSerTrpSerProLysValLeuGly***GlnAla***Al 444
QY 4830 CAGCTGCAATGGGACTCTTGATTTCTTTAAGGAAACTTGGGTAC----- 4875
Db 444 aThrThrProSer*****PheValPheGlyGlyPhePhePhePhePheAlaLe 464

```
QY 4875 ----- 4875
Db uPheLeuPheLeuArg**AlaLeuAlaLeuThrProArgLeuGluCysSerGlyLysPh 484
QY 4876 -----CAGAGTATTTCCACAAATCCTATTCAAAATTAGTGTATGATATGCAACAC 4928
Db eTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaLeuLeuProLeuSerProVa 504
QY 4929 TGTGCTAGGAGCCAGCAAAACAAAGAGGA-----GGAGAAATCAGTCAATTATGTGGGAAC 4982
Db 504 l**LeuGlyLeuGlnAlaTrpAlaAlaLeuProGly***PheMetGlyPhe*****Ar 524
QY 4983 ACATAGCAAGATATTTAGATCATTTTACTAGTGTAAAGAGCAGAGTACAAAATCA 5042
Db 524 gHisSerPheThrMetLeuLeuIleArgLeuValSerAsnSer***ProGlnVal----- 541
QY 5043 CACATGCATCAGTATAATCCAAATCATGTAATATGTCCTGTAGAAAGCAGTAGAGGA 5102
Db 542 -IleCys-----ProGlyLeuProLysCysTrp-----AspArgAr 555
QY 5103 TAAACACAAGAAATCTTAACAGTCATTTGTCATTAGACACTAAGTCTAATTATTATTAG 5162
Db 555 gGlu----- 556
QY 5163 ACATATGATATTTGAGATTTAAAAAATCTTTAATATTTTAAAAATTTAGAGCTCTCTAT 5222
Db 556 ----- 556
QY 5223 TTTTCCATGATTTCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTCTTTTCTTTT 5282
Db 557 ----ProProHisProAla*****LeuPheLeuGlyValPhePhePhePheLe 575
QY 5283 T-----TTTTTTTTTGAGATGGAGTTTGTGGTCTTTGTCCTCATGCTGGAGTGAATGG 5336
Db 575 uLeuCysPheCysPhe***AspArgProLeuLeu***HisProGlyTrpSerAlaValA 595
QY 5337 CATGAYCATAGCTACTGCAACCTCCACCTCGGGTTCAAGCAAGAGCTGTCCGCTCAGC 5396
Db 595 la-SerPheCysSerLeuGlnProGlnProGlyPheLysArgPheSerCysLeuCys 614
QY 5397 CTCCTGGGTAGATGGATTTACAGCGCCGCCACACACACACCTGCGCTAATGTTGTATTTT 5456
Db 615 LeuProCysSerTrpAspTrpArgHisGlyProPheLeuAlaAsnLeuCysIlePhe 634
QY 5457 AGTAGAGATGGGTTTCACCATCTTGGCCAGGCTGTCTCAAACTCCTGACCTCAGA 5513
Db 635 AsnArgAspThrValSerProCys***SerGlyTrpSerGlnThrProAspLeuLys 653
```

RESULT 4

C40201
artifact-warning sequence (translated ALU class C) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: C40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: C40201
A:Molecule type: DNA
A:Residues: 1-613 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; PMID:92241891; PMID:1572661
C:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

```
Pred. No.: 8.64e-34 Length: 613
Score: 407.50 Matches: 181
```

```
Percent Similarity: 33.92% Conservative: 52
Best Local Similarity: 26.35% Mismatches: 221
Query Match: 4.23% Indels: 234
DB: 4 Gaps: 17

US-09-966-880a-9 (1-5514) x C40201 (1-613)

QY 5510 GAGGTGAGGAGTTGTAGAGCAGCTGGCCAAACATGGTGAACCCCATCTCTACTAAAAAT 5451
Db 125 GluAlaArgSerArgProAlaTrpProThrTrpGlnAsnProIleSerThrLysAsn 144
QY 5450 ACAACATTTAGCCGAGTGTGGTGGCGCCTGTAAATCCATCTACCCGGGAGGTGAG 5391
Db 145 ThrLysAsn***LeuGlyMetValArgAlaProValValProArgThrGlnLysAlaGlu 164
QY 5390 GGCACAGCTTGTCTTCAACCCAGGAGGTGGAGGTTCAGTGAGCTATGTCATGCCATTC 5331
Db 165 AlaGlyGluTrpHisLysProGlyArgTrpSerLeuGln***AlaGluIleSerProLeu 184
QY 5330 CACTCCAGCATGGGCAACAGACCAAACTCCATCTCAAAAAA*****AAAAA 5271
Db 185 HisSerSerLeuGlyAsp-ArgValArgLeuLysLysLysLysLysLysLys** 204
QY 5270 AAGAAAGAGTAATCTGATGATTCATTTGCAAACTTGAATCTATGTAATAATAGAAGCTC 5211
Db 204 *****ArgGlyAlaValAla----- 211
QY 5210 TAAATTTTAAATAATTAAAGATTTTTTAAATCTCAAAATATCATAGTGTCTAATAANTA 5151
Db 211 ----- 211
QY 5150 ATTAGACTTAGTCTCTAATGACAATGACTGCTTAAGATTCTTGTGTTTATTTCCTCTAGTCT 5091
Db 212 -----HisAlaCysAsnProSerThrLeuG1 220
QY 5090 TTCTACAGGCACATATTTACATGATTTGGATTACTGATTGCTGATGTGATTTGTACT 5031
Db 220 yGlyLysGlyGlyTrpIleMetArgProGlyValArgAsp----- 233
QY 5030 CTGCTGCTTTTTTAACTAGTCAAAATGATCTAAATATCTTGCTATGTTGTCTCCACATAA 4971
Db 233 ----- 233
QY 4970 TGACTGATTTCTCCTCTCTCTTCTTTTCTGGCTCCTAGCACAGTGTCTTGCATATCATAA 4911
Db 233 ----- 233
QY 4910 GCCTAATTTGAATAGCATTTGTGAAATATCTCTGTGTAACCCCAAGTTTCCTTTAAAGAAAT 4851
Db 234 -----GlnProGlyGlnHisGlyLysThrProPheLeuLeuLysIleG1 248
QY 4850 CAAGAGTCCCATTTGCAGCTGCAGACAGCAACACAGCTGTACAGCTGTTGATATGAA 4791
Db 248 nLysIleSerTrpAlaTrpCys-----GlyArgLeu***SerHisValArgArgAr 265
QY 4790 AGCACAA-----GTGATGCTGCTGATTTTGTCCCTAGTCTTTTCCA---GAAAT 4746
Db 265 gLeuArgGlnGluAsnGlyIleAsnProGlyGlyGlyAlaCysSerGluProArgSerAr 285
QY 4745 GCATTTATCCAATTTGTGGAGCTAGTGTCTGCACCTAATGGTACGGGGCAGAAGTGCAGA 4686
Db 285 gHisCysThrProAlaTrp----- 291
QY 4685 GAATTTGCCGAAGTAAAGGCTCAGGAATTTGAATAACGTTTGTGCGCAAGAGAAATTGAAA 4626
Db 292 -----ValThrGlu***AspSerValSerLysLysLysLysLys 305
QY 4625 CTTTTCTTACACATTTATTTTCTTACATTGAGGTATTAATTTACATACAGTAAACTCCAC 4566
Db 305 s*****PhePhePhePhePhePhePheGluThrGluSerHisSerValThrGlnAl 325
QY 4565 CCATTTTAAGATGCATAATTCTAGCAATATTTTTGGATGTAAACACCCCAACTGCTACCA 4506
Db 4565 ----- 4506
```

```
Db 325 aGlyVal----- 327
Qy 4505 CAGTCAGATACAGAACCTGAATCACCACCCCTCACCCACAGAGCTTCTCCCTCTTG 4446
Db 328 ----GlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheMetProPheSe 346
Qy 4445 CAGTCCTTCC-----CCTCCCCACAGGACAAACCACTGATT 4410
Db 346 rCysLeuSerLeuLeuArgThrTrpAspTyrArgArgProHisHisAlaGlnLeu---- 364
Qy 4409 GCTTTCTTTTTCCTCCATTTTGTGGAGAAAGTTCACATGCTTGTGCCAGGCTGGTC 4350
Db 365 -----lIlePheCyslIlePheSerArgAsnGlyValLeuProCysTrpProGlyTrpSe 382
Qy 4349 TTGAACCTCGTAGTTCAGCAATCTCCACAGCTGGCCTCAAACTGCTGGATACAGG 4290
Db 382 rArgThrPro-GlyLeuMetIleHisProProLeuProProLysValLeuGlyLeuGlnA 402
Qy 4289 TGTGAGCCTCTATGCTGCGCTGATTGCTTCTGTAACATATAAGTTACTTTTCATTTC 4230
Db 402 la**AlaThrAlaProArgPhePhe----- 410
Qy 4229 AGAATTTGATGGGATACATTATCTCTCTTCATGCTGACTCTTTCATTTCAGCTCA 4170
Db 411 -----PhePhePhePhe*****LeuArgGlnSerLeuThrLeuSerProArgLeuG 429
Qy 4169 TGACTTTGGGATTATCCCAAGTAGCTTCTTCTTTGTACGACAGGTTGATTCCATT 4110
Db 429 luCysAsnGlyGluIle-----SerAlaHisCysLysLeu-----HisL 442
Qy 4109 GCATGGATATACGACATTTTGTATTATTGTTTCACTTGTGTATGGATATTGGATGCTC 4050
Db 442 euProGlyLeuCysHis----- 449
Qy 4049 TACTTTGGGTTTGTGTAATAAGTACTATCAACATTCAC---GTACCAGCTTTGTGT 3993
Db 449 roAlaSerAlaPheCysValArgGlyThrThrGlyAlaArgThrMetProSer***PheP 469
Qy 3992 GCATATACATTCAGAATTGATATTTTAATAAGTAGTAGCAGGTGACAGACCCTGACTTTA 3933
Db 469 heValPheLeuValGluMetGlyPheCysHisValGlyGlnAla----- 483
Qy 3932 AAAGTTAAATACAAATATACGCCTAATAAGTAGTCACTATAGACAAACAGCTCTTG 3873
Db 484 -----GlyLeuGluLeuLeuA 489
Qy 3872 TAGTGTCTTCATTATTTT----- 3853
Db 489 laSer***SerThrHisLeuCysLeuProLysCysTrpAspTyrArgArgGluProLeuH 509
Qy 3852 -----TAAATTTTCTTCTTATTTTGTGGCAAGGTCTGCTCTGTCACC 3807
Db 509 isProAla*****PhePhePhePhePhe***AspArgValSerLeuCysHisP 529
Qy 3806 CAGGCTGGAGTCAGTGGCGCAAGCACTCTCAGTCGACGCTTGACTCTCTGGCTCCAA 3747
Db 529 roGlyTrpSerAlaMetAlaArgSerArgLeuThrAlaSerSerThrSerArgValTyrA 549
Qy 3746 AGATATCTCTGCTCAGCCCTCTAGTAGCTGGGACCACAGCACATGCCACCTGCCGG 3687
Db 549 laIleLeuLeuProGlnProSerAlaTyrValGlyLeuGlnAlaProAlaProCysProA 569
Qy 3686 CTAATTTTGTATTAG-----AAACAGGTTTGTCCATGTGTGCTCAGGCTGGCTGA 3634
Db 569 laAsnPheLeuTyrPhe*****LysTrpGlyPheAlaMetLeuAlaArgLeuValIserA 589
Qy 3633 ACTCCAGGCTCAAGCGATCCACCCACCTCAGCTCCGCTCCAGTGTGGGATATACAGGCT 3574
Db 589 snSer--TrpProHisAspProThrPheAlaSerGlnSerAlaGlyIleThrGlyVa 608
Qy 3573 GAGCCACCAACC 3561
Db 608 lSerHisCysThr 612
```

```
RESULT 5
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of
```

```
Alignment Scores:
Pred. NO.: 6,35e-27 Length: 597
Score: 342.50 Matches: 196
Percent Similarity: 33.29% Conservative: 60
Best Local Similarity: 25.49% Mismatches: 209
Query Match: 3.55% Indels: 305
DB: 4 Gaps: 29

US-09-966-880A-9 (1-5514) x E40201 (1-597)
Qy 3559 CAGGTGTGTGCTCAGGCTGTAATCCACGACCTTGGGAGGCTGAGTGGGTGG-ATC 3617
Db 2 GlnValTrp**LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArgIle 21
Qy 3618 GCTTGAGCCCTCGAGTTCGAAGACCAGCCCTGACACATCGGCAAAACCCCTGTTCTATATAC 3677
Db 22 Thr**GlyTrpGluPheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr--** 41
Qy 3678 A-----AAAATGCCGGCATGGTCATGTGCTGTGTCGCCAGCTACTAGGGGGCT 3731
Db 41 *LysTyrLysIleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgGluAl 61
Qy 3732 GAGGCAG---GAGAATCTTTGGAGCCAGGAGTCAAGGCTGCATGCATGACGCTGCTGG 3788
Db 61 aGluThrGlyGluAsnCysLeuAsnProGlySerLysValCysGlyGlu--lIleVal--- 79
Qy 3789 CCAGTGCATCCAGCCTGGGTGACAGGACCAGA---CCTTGGCTCAAAAAATAAGAAGA 3845
Db 80 ProSerTyrSerGlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLys***** 99
Qy 3846 AAAATTAATAAATAATGGAACAACACTACAAAGAGCTGTTGTCTCTAGCTGAGCTACTAGT 3905
Db 100 ***AlaArg----- 102
Qy 3906 TAGGCTGATATTTTGGTATTATTAACCTTTAAAGTCAAGGCTGTGTCACTGCACCTACATTAT 3965
Db 102 ----- 102
Qy 3966 TAAATATCAATTCATCAATGTATATCCACAAAGACTGGTAGCTGAATCTTCATAGTAC 4025
Db 103 -----CysGlySerSerCysLeu***SerGlnArg----- 112
Qy 4026 CTTTATTCACAAACCCCAAGTAGAGACTATCCAATATCCATCAACAGTGAACAAAT 4085
Db 113 ---PheGlySerSerArgArg----- 118
Qy 4086 AAACAAAATGTCTATATCCATGCAATGGAATACCAACCCCTGCAAGTACAAGGAGAGCT 4145
Db 119 -----GlyGlySerProGluValGlySerLeuArgProAla 130
```


A:Reference number: A40201

A: Accession: E40201

A:Molecule type: DNA

A:Residues: 1-597 <CL.A>

R: Claverie, J. M.

Genomics 12, 838-841, 1992

Genomics 12, 638-641, 1992

A: Reference number: A40200: MULT: 92241891: PMTD: 1572661

A:Reference Number: A40.

```

1011PRGNUM: and
1012X,concent:
1013C:Comment: This "warning" entry is a consecutive translation for all 6 words; a function of

```

c;comment: This warning entry is a conceptual translation in all 6 reading frames or in-frame stop codons are shown as /v/

III-frame stop codons are shown as 'X'.

Alignment Scores:

Pred. No.:	4, 4.9e-25	Length:	597
Score:	325.00	Matches:	198
Percent Similarity:	33.12%	Conservative:	56
Best Local Similarity:	25.81%	Mismatches:	211
Query Match:	3.38%	Indels:	303
		Gaps:	30

US-09-966-880A-9 (1-5514) x E40201 (1-597)

Qy	5508	GGTCAGGAGTTTGAGACCGAGCTGGCCAAACATGGTGAAGACCCCATCTCTACTAAAAATAC	5449
Db	24	GlyTrpLupheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr**LysTyr	43
Qy	5448	AAACATTAGCCGAGTGTGGTGGGGCCCTGTAATCCCATCTACCCGGAGAGCTGAG--	5391
Db	44	Lys-IleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgGluAlaGlu	63
Qy	5390	-GGACACAGCTTGGCTTCAACCCAGAGGTGGAGGTGCAGTGCAGTATGRTCATGCCATT	5332
Db	63	rGly-GluAsnCysLeuAsnProGlySerLysValCysGlyGlu--Ile---valProSe	81
Qy	5331	CCACTCCAGCATGGGCAACAGACCAAACTCCCATCTCAAAAAAANAANAANAANA	5272
Db	81	rTyrSerGlyLeuGlyAsnLysSerLysThrProSerGln-LysLysLys*****A	101
Qy	5271	AAAGAAAGAGTAATACCTTGATCATCTGTCAAACTTGATCTATCGGAAAAATAGACAGCT	5212
Db	101	laArg-----	102
Qy	5211	CTAAATTTTAAATATTTAAAGATTTTTTAAATCTCAAATATCATAGTGTCTAATAATAAT	5152
Db	102	-----	102
Qy	5151	AATTAGACTTAGTGCTCAATGACAATGACTGTTTAAGATTCTTGTGTT-----T	5104
Db	103	-----CysGlySerSerCysLeu**SerGlnArgp	113
Qy	5103	ATTCCCTCTAGTCTTCTACAGGCACATATTTACATGATTTGGATTGATGATGATGT	5044
Db	113	heGlySerSerArgArgGlyGly-----	120
Qy	5043	GTGATTTTGTACTGTCTGCTTTTTTAACTAGTCAAAATGATCAATATCTTGCATGT	4984
Db	121	-----SerProGluValGlySerL	127
Qy	4983	TGTTCCACATATGACTGATTTCTCCTCTCTTTGTTTCTGGCTCCTAGCACAGTGTC	4924
Db	127	euArgProAla**Pro-----	132
Qy	4923	TTGATATCATAGCACTAATTTGAATAGCATTTGTGGAATACTCTGTGTAACCCAGTT	4864
Db	133	-----ThrTrpArg-----AsnProIles	139
Qy	4863	TCCTTAAGAAATCAAG-----AGTCCCCCATTGCA	4834
Db	139	erThrLysAsnIleLysLeuAlaGlyArgGlyAlaCysLeu***SerGlnLeuLeuG	159
Qy	4833	GCTGCACAGCAACACAAGCTGTACAGCGTGGTTGATATGAAGACACAGTATGCCTG	4774

```
Db 471 ----- 471
QY 3951 GTGACAGACCCCTGACTTTAAAGTTAAATACCAAAATATACAGCCTAACTAAGTAGCTCAT 3892
Db 472 ----- 472
QY 3891 CTAGGACAACAGCTCTTTGTAGTTGTTCCATTTATTTTAAATTTTCTTCTTATTTT 3832
Db 473 ----- 473
QY 3831 TGAGGCAAGCTCTGGTCTGTCCAGCCTGGAGTGCGAGCGCGCAAGCA----- 3781
Db 474 ----- 474
QY 3780 -----CTGCTCAGTCAGCCTTGACCTCTCGGCTCCAAAGATTCCTC---CTGCTCAG 3730
Db 475 ----- 475
QY 3729 CCCCCTAGTAGCTGGGACACACATGCCACCATGCCGCGCTAATTTTGT----- 3675
Db 476 ----- 476
QY 3674 -ATAGAAACAGGTTTGGCAGTGTCTGCTCAGCTGTGTGAAGTCCAGGCTCAAGCGA 3616
Db 477 ----- 477
QY 3615 TCCACCCAC-CTCAGCCTCCAGAGTGGGATTACAGCGTACAGCCACACACCTGCG 3557
Db 478 ----- 478
QY 3556 T 3556
Db 479 ----- 479
QY 597 a 597
Db 480 ----- 480
RESULT 8
A:Title: warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <LA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of a
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the
Alignment Scores:
Pred. No.: 6 52e-25 Length: 627
Score: 323.50 Matches: 172
Percent Similarity: 31.50% Conservative: 57
Best Local Similarity: 23.66% Mismatches: 191
Query Match: 3.36% Indels: 310
DB: 4 Gaps: 19
US-09-966-880A-9 (1-5514) x A40201 (1-627)
QY 5509 AGTCTAGAGTTGTGAGACGAGCTGCCCAACATGGTGAACCCCATCTCTACTAAAAATA 5450
Db 5510 ----- 5510
QY 5449 CAAACATTAGCCGAGTGTGGTGGCGCTGTATCCCATCTACCCGGAGCGTGGAG 5390
Db 5450 ----- 5450
QY 42 GlnLysSerArgAla***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArg 61
```

```
QY 5389 CGACAGCTTTGCTTGAACCCAGGAGGTGGAGTTGCAGTCAGTCAGTCATGTCATGCCATTC 5330
Db 5390 ----- 5390
QY 5329 ACTCCAGCATGGCAACAAAGCAATCCATCTCAAAAAAATAAAAAAATAAAAAA 5270
Db 5330 ----- 5330
QY 5269 AGAAGAGTAATCTGATCATTTGCTCAAACTTCAATA-----CTATGGAAAAATAG 5219
Db 5270 ----- 5270
QY 5218 AAGAGCTCTAAATTTTAAATATTAAGATTTTTTAAATCTCAAAATATCATAGTGTCTAA 5159
Db 5219 ----- 5219
QY 5158 TAATAATAATAGACTTAGTGTCTAATGACAATGACTGTGTAAAGATCTTGTGTATTATTC 5099
Db 5159 ----- 5159
QY 5098 TCTAGTCTTTCTACAGGCACATATTTACATGATTGATGATTACTGATTCATGCTGATG 5039
Db 5099 ----- 5099
QY 5038 TTTGTACTCTGCTGCTTTTAACTAGTCAAAATGATCTAAATATCTTGTCTATCTGTTTC 4979
Db 5039 ----- 4979
QY 4978 CCACATAATGACTGATTTCCTCCTCTTGTGTTTCTGCTCTAGCAGTGTCTTGCA 4919
Db 4979 ----- 4919
QY 4918 TATCAATAGCCTAATTTGNAATAGCATTTG-----TGAATATCTCTGGTAAACCAAGTTT 4863
Db 4919 ----- 4863
QY 4862 CCTTAAAGAAATCAAGATCCCATTTGCAGCTGCAGACACACAAAGCTGTACAGCGT 4803
Db 4863 ----- 4803
QY 4802 GGTGTATGAAGACACAGTGTGCTGCTCACTAATGGTACGGGAGAGTGCAGAGAA 4743
Db 4803 ----- 4743
QY 4742 TTTATCCAATTTCTGGAGCTAGTGTCTCACTAATGGTACGGGAGAGTGCAGAGAA 4683
Db 4743 ----- 4683
QY 4682 T-----GlnHisPheGlyArgProArgAlaAspHI 230
Db 4683 ----- 230
QY 4667 GCTCAGGAATTTCAATTAACGTTTGC----- 4643
Db 4668 ----- 4643
QY 4642 -----TGCACAGAG----- 4634
Db 4643 ----- 4634
QY 4633 -----AATGAACTTTCTTACACATTTATTTTCTTACATTGAGGTATATTTTACA 4581
Db 4634 ----- 4581
QY 4580 TACAGTAAACTCCACCCATTTTAAGATGCATAATTTCTACGAATATTTTGGATGTAACAC 4521
Db 4581 ----- 4521
QY 4521 -----TrpArgAspLeuGlySer 344
Db 4522 ----- 344
```



```
Db 146 ---GlySerIleSerLeu***SerGlnLeu---LeuLysArgLeuGlyGlnGluAspHis 163
QY 4278 -----
Db 164 LeuThrTrpGluValLysAlaGln**AlaMetIleAlaProLeuTyrSerSerLeuGly 183
QY 4279 -----AGAGGCTCACACCTGTAATCC 4299
Db 184 AsnIleProArgProCysLeuLys*****ProSerMetAlaAlaArgLeu***Ser 203
QY 4300 CAGCACTTTGAGAGCCACGTCGGAAGATTGCTAGAACTCAGGAGTTCAAGACCAGCGCTG 4359
Db 204 GlnHisPhe-ArgSerHisGlyArgCysMetLeuGlnLeuArgSerSerArgThrAlaTr 223
QY 4360 GGCACACAGTGAACCTCATTCTCCACAAAAATGGGAAAAAGAAAGCAAAATCAGTGG 4419
Db 223 pAlaThr-CysArgAsnProThrSerThrLysLysArgTyrLysAsn***ProAspLeuV 243
QY 4420 TTGTCCTGTGGGAGGGAAGGACTGCAAGAGGGAAGAAGCTCTGTGGGGTGAGGGTG 4479
Db 243 alAla***AlaCysSerProSerTyrSerLysGly-----TrpGly-Arg--- 257
QY 4480 GTGATTCAAGTTCTGTATCCTGACTGTGTAGCAGATTGCGGTGCTTTACATCCAAAAATA 4539
Db 258 -----ArgIleThr***Pro-----GlyLysSer---ArgLeuSerGluPro***Leu 272
QY 4540 TTCGTAGAATTATGCATCTTAATAGGTGGTGAGTTTACTGTATGTAATATTACCTCAATG 4599
Db 273 -----HisLeuCys---ThrProAlaTrpAla----- 280
QY 4600 TAAGAAAAATAATGTGTAGAAAAAGTTTCAATCTCTGCCAGCAAAACGTTATTCAAT 4659
Db 281 -----ThrTyrGlnAspProAla***Lys***** 289
QY 4660 TCCTGAGCCCTTTACTTCGCAAAATCTCTGCACCTTCTGCCCTTACCATTAGTGACAGC 4719
Db 290 -----*****PheSerGlyArgValLeuValCysCysPro-----GlyTrpSer 304
QY 4720 ACTAGTCCCAAAATTTGGATA---AATGCATTTCTGGAAGACTAGGGAC----- 4767
Db 305 ThrGluValGlnSerTrpGluThrGluPro***LeuProArgSerSerAspProProAla 324
QY 4768 -----AAAATCCAGGCATCACTTGTGCTTTTCATATCAACCA----- 4803
Db 325 ProAlaPhe***ValAlaGlyThrThrGlyLeuCystyrGlnIleTrpLeuIlePheVal 344
QY 4804 -----CGCTGTACAGCTTGTGTGTCTGTCTGCAGCTGCAATGGGACTCTTGATTCTTT 4857
Db 345 SerPhePheCysArgGlyGlyValSerThrCys---CysProGlyCysSer----- 360
QY 4858 TAAGGAACTTGGTTACCAGAGTATTTCCACAAATGCTATTCAAAATAGTGTATTATGAT 4917
Db 361 -----*****ThrProGlu 364
QY 4918 ATGCAAGACACTGTCTAGGAGCCAGAAACAAAGAGGAGAGAA---ATCAGTCAATTAT 4974
Db 365 LeuGluHisThrProAlaMetAlaProLysValLeuGlyLeuGlnAlaCysSerHis-Al 384
QY 4975 GTGGGAACACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAAAAGCAGCAGAGTA 5034
Db 384 atrp*****Phe-----GlnAlaGlySerTr 394
QY 5035 CAAAATCACACATGCAATCAGTATATCAAAATCATGTAATATATGTCCTGTAGAAAGAC 5094
Db 394 pTyrValAlaGlnAla-GlyVal-----GlnArgCysAsnHis----- 406
QY 5095 TAGAGGAATAAACACAAGAATCTTAACAGCTATTGTCAATTAGAC----- 5138
Db 407 -----GlySerLeuSerLeuAspPheProGlyGlnValI 418
QY 5139 -----ACTAAGTCTCAATT 5151
Db 418 leuLeuLeuProGlnProPheGlu***LeuGlyLeuGlnAlaTyrAlaThrArgSerGly* 438
```

```
QY 5152 ATTATTATTAGACACTATGATATTGAGATTTAAAAAATCTTTAATATTTTAAAAATTTAG 5211
Db 438 **Phe-----LeuT 441
QY 5212 AGCTCTCTCTATTTCCTCATAGTATTCAAGTTTGCATGATCAAGATATTACTCTTTCTTT 5271
Db 441 yrLeuPhePheValGluValGlyPheArgHisValAlaGlnAlaValLeuGluLeuLeu 461
QY 5272 TTTTCTTTTCTTTTCTTTT----- 5288
Db 461 erTrpSerIleHisLeuProTrpIleLeuLysCysTrpAspTyrArgArgAlaAlaMetL 481
QY 5289 -----TTTTTTCAGATGGAGTTTGTCTTGTTCCTCATGCTGGAGTGAATGGC 5337
Db 481 euGly*****PheArgGlnGlyLeuGly-MetLeuProArgLeuGluTyrArgGly 500
QY 5338 ATGAYCATAGCTCACTCAACCTCCACCTCCTGGTTTCAAGCAAGCTGTGCGCTCAGCC 5397
Db 501 AlaIleMetAlaHis***AlaLeu-ThrSerGlnValLys***SerSerCysPro-SerL 520
QY 5398 TCCCGGCTAGATGGGATTACAGCGCCACCACACACTCGGCTCTCAAACTCC 5503
Db 520 euLeuSerSerTrpAspTyrArgLeuMetLeuProAspLeuAlaAsnPheCysIlePheP 540
QY 5458 GTAG-----AGATGGGGTTTCCCATGTTGGCCAGGCTGGTCTCAAACTCC 5503
Db 540 heLeu***ArgTrpGlyPheAspMetLeuProArgLeuPheLeuAsnSer 556
```

RESULT 10

S41044

chromosomal protein - human

C:Species: Homo sapiens (man)

C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000

C:Accession: S41044

R:Yeo, J.P.; Alderuccio, F.; Toh, B.H.

Nature 367, 288-291, 1994

A:Title: A new chromosomal protein essential for mitotic spindle assembly.

A:Reference number: S41044; MUID:94166884; PMID:8121495

A:Accession: S41044

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-418 <YEO>

A:Cross-references: GB:D26953; NID:g537529; PIDN:AAB68050.1; PID:g537530

C:Superfamily: human 48.2K chromosomal protein

C:Keywords: chromosomal protein

Alignment Scores:

Pred. No.:	1.97e-14	Length:	418
Score:	224.00	Matches:	91
Percent Similarity:	37.39%	Conservative:	32
Best Local Similarity:	27.66%	Mismatches:	74
Query Match:	2.32%	Indels:	132
DB:	2	Gaps:	17

US-09-966-880A-9 (1-5514) x S41044 (1-418)

```
QY 3550 ACAGACAGCCAGGTG---TGG-----TGGCTCACGCCTGTAATCCACGACCTCTGGAG 3600
Db 163 ThrGlnSerTyrValAspTrpGlyArgTrpLeu-----IleTrpGlu 176
QY 3601 GCTGAGGTGGGTGATCGCTTGAGCCCTTGAGCCCTGAGCTCAAGACCCAGCATGAGCAAC 3660
Db 177 AlaLysAlaGlyGluSerLeuGluValArgSerArgProAlaSerGlnSerArgArg 196
QY 3661 AACCCCTGTTTCATA-----ACAAAAAT 3684
Db 197 AsnSerValSerThrLysAsnIleLysIleSerProValSerThrLysAsnIleLysIle 216
QY 3685 AGCCGGCATGGTGG-----CATGTGCTGTGTGCTCCAGCTACTAGGGGG 3729
Db 217 SerGlnThrTrpTyrLeuPheGlyGlyValHisLeu---LeuValProThrThrArgAsp 235
```



```
Db 263 yLysSerArgLeuSerGluPro***LeuHisLeuCysThrProAlaTrpAla----- 280
QY 4715 TCACCTAAGTACGGGGCAGAAAGTCACAGAAATTTTCGGAAGTAAAGGGCTCAGGAATTT 4656
Db 281 :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 281 -----ThrTyrGlnAspProAla***Lys*****PheSerGly---ArgValLe 297
QY 4655 GAATAACGTTTGC-----TGGCAAGAGAATTGAAACTTTTCTTACACATATATTTT 4605
Db 297 u-----ValCysCysProGlyTrp----- 303
QY 4604 TCTTACATTGAGGTATAATTTACATACAGTAAACTCCACCATTTCAGCATATTTCT 4545
Db 304 :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 304 -----SerThrGluValGln----- 308
QY 4544 ACGAATATTTTGGATGTAACACCCCAAACTGCTACACAGTCAGGATACAGAACTGA 4485
Db 309 :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 309 -----SerTripleuThrGluPro** 315
QY 4484 ATCACCACTCTACCCACCACAGAGCTTCTCCCTCTTTGAGTCCTCCCTCCCCACAG 4425
Db 315 *LeuProArgSerSerAsp-----ProProAlaProAlaPhe** 328
QY 4424 GACA-----ACCACGATTTG-----CTTCTCTTTTTCCTCCATTTTGT 4386
Db 328 *ValAlaGlyThrThrGlyLeuCysTyrGlnIleTrpLeuIlePheValSerPhePheC 348
QY 4385 GGAGATGAGGTTTCACTGTTGTCGCCAGCGTGTCTTGAACCTCCAGCTTCTAGCAATC 4326
Db 348 sArgGlyGlyValSerThrCysCysProGlyCysSer***ThrProGluLeu----- 365
QY 4325 TTCCCACGTCGCTCTCAAGAGCTGGGATTACAGGTGTGAGCCCTATGCTCCCTGCTGA 4266
Db 366 -GluHisThrProAlaMetAlaProLysValLeuGly-----LeuGlnAlaCys----- 381
QY 4265 TTTGCTTTCTGTAAGTACTATTTCTATTTCTAGAAATTTGCTATGTCATACATAT 4206
Db 382 -----SerHisAlaTrp*****PheGlnAlaGly----- 392
QY 4205 CTCCTCTTCATGTCGACTCTTTTCATTTAGCGTCAGCATTTGGG----- 4160
Db 393 -SerTrpTyrValAlaGlnAlaGlyValGlnArgCysAsnHisGlySerLeuSerLeuAs 412
QY 4159 -----ATTCATCCCAAGTAGCTTCTCTCTTTGTACTCAGCGGTGG 4119
Db 412 pPheProGlyGlnValIleLeuLeuProGlnProPheGlu***LeuGlyLeuGlnAlaTy 432
QY 4118 TATTCATGTCANGATATAGCACATTTTGTATTTGTTCACCTTTGTTGATGGATATTG 4059
Db 432 rAlaThrArgSerGly***PheLeuTyrLeuPhePheValGlu----- 446
QY 4058 GATAGTCCTACTTTGGGGTTTGTGAATAAAGGTACTATGAAACATTCACGTACAGTCT 3999
Db 447 -----ValGlyPhe-----ArgHisValAlaGlnAlaValLeuGluLe 459
QY 3998 TTGTGTGGATATACATTGAGAATTGATATTTTAATAATGATGTCAGGTGACAGACCCTG 3939
Db 459 u----- 459
QY 3938 ACTTTAAAGTTAAATACAAAATATCAGCCTAACTAAGTACTATCTAGTACACACAGC 3879
Db 460 -----LeuSerTrpSerIleHisLeuProTr 468
QY 3878 TCTTTGTAGTCTTCTCCATTATTTTAAATTTTCTCTTT-----ATTTTTCAGGCA 3825
Db 468 pleuLeuLysCysTrpAspTyrArgArgAlaAlaMetLeuGly*****PheArgI 488
QY 3824 AGGTCTGTGTCCTCCACCGAGCTCGAGTGCAGTGCAGCAAGCACTGCTCAGTCAGCGCT 3765
Db 488 nGlyLeuGlyMetLeuProArgLeuGluTyrArgGlyAlaIleMetAlaHis-***AlaL 508
QY 3764 TGACCTCTCGGCTCCAAAGATTCCTCGCTCAGCCCTAGTACCTGGGACCAAGGC 3705
Db 508 euthrSerGlnValLys***SerSerCysProSerLeuLeuSerSerTrpAspTyrArgL 528
```

```
QY 3704 ACATGCCACCATCCCGCTAAATTTTGTATTATAG-----AAACAGGTTTCCCA 3655
Db 528 euMetLeuProAspLeuAlaAsnPheCysIlePhePheLeu***ArgTrpGlyPheAspM 548
QY 3654 TGTTCCTCAGCGCTGGTCTTGAACCTCCAGGCTCAAGCGATCCACCCACCTCAGCCTCCCA 3595
Db 548 etLeuProArgLeuPheLeuLeuAsnSer***AlaGlyAlaTyr--ThrCysHisGlySer** 567
QY 3594 GAGTCTGGGATTACAGCGCTG 3573
Db 567 *SerAlaGlyIleThrGlyVal 574

RESULT 12
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46010
R:Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A:Title: Identification and partial characterization of a candidate gene for X-linked
A:Reference number: A46010; MUID:93224131; PMID:8468040
A:Accession: A46010
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-100 <WON>
A:Cross-references: GB:S58722; NID:g299470; PIDN:AAB26149,1; PID:g299471
A:Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:
Pred. No.: 1,94e-13 Length: 100
Score: 213.50 Matches: 49
Percent Similarity: 71.95% Conservative: 10
Best Local Similarity: 59.76% Mismatches: 20
Query Match: 2.21% Indels: 3
DB: Gaps: 1

US-09-966-880A-9 (1-5514) x A46010 (1-100)

QY 5276 TTTTTTTTTTTTTTTTTTTCAGATGGAGTTTGTGCTTGTGCTCCATGCTGGAGTGAATG 5335
Db 1 PhePhePhePhePhePheGluThrGluSerCysSer-ValAlaGluAlaGlyValGlnTr 20
QY 5336 GCATGAYCATAGTACTACTGCAACCTCCACTCTCTGGTTTCAAGCAAGCTGTGCGCTCAG 5395
Db 20 pCysAspLeuGlySerLeuLysSerProProGlySerSerAspSerProAlaSerA 40
QY 5396 CCTCCCGGTAGATGGGATTACAGCGCCGCCACACACACTCGGCTAATGTTGTATTTT 5455
Db 40 laSerArgValAlaGlyIleThrGlyMetHisHisThrGlnLeuIlePheValPheL 60
QY 5456 TAGTAGATAGTGGGTTTTCAC---CATGTTGGCCAGGCTGTCTCAAACTCTCGACCTCAG 5512
Db 60 euValGluThrGlySerHisMetGlnLeuSerAspSerThrLeuValIleThrThrAlaG 80
QY 5513 AG 5514
Db 80 ln 80

RESULT 13
S41044
chromosomal protein - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: S41044
R:Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A:Title: A new chromosomal protein essential for mitotic spindle assembly.
A:Reference number: S41044; MUID:94166884; PMID:8121495
A:Accession: S41044
A:Status: preliminary
A:Molecule type: mRNA
```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Db	2	ArgAlaTrpTrpLeuThrSerValIleProAlaLeuTrpAspAlaGluValGlyGlyLeu	21
Qy	3619	CTTGAGCCCTGGAGTTCAAGACGAGCCTGAGCAACATGCGCAAAACCTGTTCTTATATAACA	3678
Db	22	LeuGluThrArgSerSerArgProAlaTrpAlaThr***GlnAspProSerSerIle-Ly	41
Qy	3679	A-----AAATTAGCCGGGCATGGTGGCATGTGCCTGTGGTGC	3714
Db	41	sIleLys***Asn***LeuGly*****AlaGlyHisGlyGly***HisLeu***Se	61
Qy	3715	CCAGCTACTAGGGGCTGAGCAGGAGAACTTTTGGAGCCGAGGAGTCAAGGCTGCACCT	3774
Db	61	rGln-----HisPheGlyMetLeuLys***Glu-----	70
Qy	3775	GAGCAGTGCTTGGGCCACTGCATCCAGCTGGGTGACAGGACCAGACCTTGGCTCAAAA	3834
Db	71	-----AspCysLeuArgProGlyValGlnAspGln-----	80
Qy	3835	AAATAAGAGAAAAATTTAAAAATAATATGGAACAACACTACAAAGAGCTGTGTCTAGATG	3894
Db	80	-----	80
Qy	3895	AGCTACTTAGTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACTG	3954
Db	81	-----LeuGly-----	82
Qy	3955	CACCTACATTATAAATATCAATTCATATGATATCCACACAAGAGACTGGTACGTGAAT	4014
Db	83	-----GlnHisSerLysThr-----	87
Qy	4015	GTTCATAGTACCTTTATTCACAAAACCCAAAGTAGAGACTATCCAAATATCCATCAACA	4074
Db	88	-----ProAlaLeu***Lys-----	92
Qy	4075	AGTGAACAATAACAAAAATGTGCTATATCCATGCAATGGAATACCACTGCAGTACAA	4134
Db	93	***AsnLys-----IleSerTrpVal*****Gln-----	103
Qy	4135	AGGAAGAAGCTACTTGGGGATGAATCCCAAGCTATGACGCTAAATGAAAGAGTCAACA	4194
Db	103	-----	103
Qy	4195	TGAAGGAGAGATAATGTATGCCATACGAAATCTAGAAAAATGAAAGTAACTTATAGTTA	4254
Db	103	-----	103
Qy	4255	CAGAAACAAATCAGGCCAGCATAGAGGCTCACACCTGTAAATCCCAAGCACTTTGAGAGG	4314
Db	104	-----GlyMetValAlaAsnIleCysAsnProSerThrLeuGlyCy	117
Qy	4315	CCACGTGGGAAGATTCTAGAACTCAGGAGTTCACAGCAGGCTGGGCAACACAGTAA	4373
Db	117	s***SerArgIleAla***AspGlnGluPheLysThrSerLeuGlyAsnIleAlaAr	137
Qy	4374	ACTCCATTCTCCACAAAAATGGGAAAAAAGCAAAATCAGTGGTGTGCTGTGGGGA	4433
Db	137	g-----ProGlnLeuTrpLysAsnLys-----	144
Qy	4434	GGGGAAGCACTGCMAAGAGGAGAGCACTCTGGTGGGTGAGGTTGGTGTATTCAGGTTCT	4493
Db	145	-----IleLysLeu-----	147
Qy	4494	GTATCCTGACTGTGGTAGCAGTTTGGGGTGTATACATCCAAAAATATTCGTAGAATTATG	4553
Db	148	-----AlaGly*****HisProAlaAsnPheIle-----	159
Qy	4554	CATCTTAAATGGGTGGAGTTTACTGTATGTAATATATACCTCAATGAAGAAAAATAAT	4613
Db	159	uPheLeu***SerTrpGlyLeu-----	166
Qy	4614	GTGPAAGAAAGTTTCAATTCTCTCTGCCAGCAACAGTTATTCAAATTCCTGAGCCCTTTA	4673
Db	167	-----AlaMetLeuProLysLeuValLeuAsnSer-----	176

```

A:Molecule type: DNA
A:Residues: 1-196 <RES>
A:Cross-references: EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PID:g288145

Alignment Scores:
Pred. No.:      4,11e-11
Score:          192.00
Percent Similarity: 65.88%
Best Local Similarity: 54.12%
Query Match:      1.99%
DB:              2

US-09-966-880A-9 (1-5514) x 138022 (1-196)

Qy      5512 CTGAGTCAAGAGTTTGAGACCAGCCTGGCCACATGGTCAAAACCCACTCTCTACTAAAA 5453
Db      4   LeuArgSerGlyValGlnAsnTyrProGlyGlnHisGlyLysIleProSerLeuLeuLys 23

Qy      5452 ATACAAACATTAGCCGAGTCTGGTGTGTGGCGCCTGTAAATCCCATCTACCCGGGAGCGTG 5393
Db      24   IleGlnLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArgLeu 43

Qy      5392 AGGGACAGCTTTCCTTGAACCCAGGAGGTGGAGGTGCAGTGCAGTATGRTCATGCCAT 5333
Db      44   ArgGlnGlnAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHisLeu 63

Qy      5332 TCCACTCCAGTCCGCGCAACAGACCAAACTCCATCTCAAAAAAANAANAANAANA 5273
Db      64   CysIleProAlaTrpValThr-GluGlyAspSerValSerLysGlnAsnLysThrLysAs 83

Qy      5272 AAAAGAAGAGATA 5260
Db      83   nGluGlnHisLeu 87

```

Search completed: June 14, 2003, 18:57:31
Job time : 214.707 secs

QY 4674 CTTGCGAAATTTCTCTGCACCTTCTGCCCCGTACCATTTAGGTGACAGCACTAGCTCCACAA 4733

Db 176 ----- 176

QY 4734 TTGGTAATATGCATTTCTTGAAAAGACTAGGGACAAAATCCAGGCATCATTGTGTCTTC 4793

Db 177 -TrpSerGlnAlaIleLeuLeuGlnHisProLysVal----- 189

QY 4794 ATATCAACCACGCTGTACAGCTTGTTGCTGTCTGCAGCTGCAATGGGCACTCTTGATT 4853

Db 189 ----- 189

QY 4854 TCTTTAAGGAACTTTGGTTTACCAGAGTATTTCCACAAATGCTATTCAAATATTAGTGCTTA 4913

Db 190 -----LeuGlyLeuGlnMetLeuAlaThr----- 197

QY 4914 TGATATGAAGACACTGTGCTAGGAGCCAGAAAACAAGAGGAGAGAAATCAGTCATTA 4973

Db 197 ----- 197

QY 4974 TGTGGGAACAACATAGCAAGATATTTTATGATCATTTTGACTAGTTTAAAAAAGCAGCAGAGT 5033

Db 197 ----- 197

QY 5034 ACAAATCACATGCAATCAGTATATATCCAAATCATGTAATATGTGCCTGTAGAAAGA 5093

Db 197 ----- 197

QY 5094 CTAGAGGAATAAACACACAAGAACTTAAACAGTCATTTGTTCATTAGACACTAAGCTCAATTAT 5153

Db 198 -----MetProCys*****ThrGlnLeuIleLe 208

QY 5154 TATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTTAAATTTTAAAAATTTAGAG 5213

Db 208 u----- 208

QY 5214 CTCCTTATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGTATTACTCTTCTTTT 5273

Db 208 ----- 208

QY 5274 TTTTCTTTTCTTTTCTTTTGTGATGAGAGTCTTTGGTCTGTGTGCCCATGCTGAGAGTGGA 5333

Db 209 -----PheTyrPheTyrArgAlaGlyValLeuLeu-CysCysProSerTrpSer---- 224

QY 5334 TGGCATGATCAGCTCAGCTGCAACCTCCACCTCGGTTCAGCAAGCTGTCGCCTC 5393

Db 225 -----***Thr-ProGlyLeuLysGlnSerSerTyrPhe 235

QY 5394 AGCCTCCCGGTAGATGGATTACAGCGCCACCA-----CAC 5435

Db 236 SerIleProLysCysTrpAspTyrArgCys**ProProCysProAla*****Pro 255

QY 5436 TCGGCTAATGTTGTATTTTATTTAGTAGATGGGGTTTCACCATGTTGCCAGGCTGGTCT 5495

Db 256 Ser***PheTyrPheIlePhe-IleGluLeuGlySerCysTyrValAlaGlnAlaGlyLe 275

QY 5496 CAAACTCCTGCACCTCA 5511

Db 275 uGluLeuLeuValSer 280

RESULT 15

I38022

hypothetical protein - human

C:Species: Homo sapiens (man)

C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999

C:Accession: I38022

R:Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zellstra-Ryalls, J.H.; Somerville, R.L.

Cancer, Res. 50, 5658-5667, 1990

A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat

A:Reference number: I38021

A:Accession: I38022

A>Status: preliminary; translated from GB/EMBL/DBJ

RESULT 15
I38022

hypothetical protein - human
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C:Accession: I38022
R:Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zeilstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A:Reference number: I38021
A:Accession: I38022
A:Status: preliminary; translated from GE/EMBL/DBDJ

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:11:13 ; Search time 69.1643 Seconds

(without alignments)
6613.252 Million cell updates/sec

Title: US-09-966-880A-9

Perfect score: 9659

Sequence: 1 acagacgaatacatagttccca.....tcaaacctcctgacctcagag 5514

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPro.spool/US09966880/runat_14062003_175524_10304/app_query.fasta_1.9493
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880_ECGN_1_196/runat_14062003_175524_10304 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	554	5.7	593	1 ALU7_HUMAN	P39194 homo sapien
2	553	5.7	591	1 ALU1_HUMAN	P39188 homo sapien
3	526	5.4	593	1 ALU6_HUMAN	P39193 homo sapien
4	522.5	5.4	591	1 ALU8_HUMAN	P39195 homo sapien
C 5	490.5	5.1	591	1 ALU1_HUMAN	P39188 homo sapien
C 6	489.5	5.1	593	1 ALU7_HUMAN	P39194 homo sapien
C 7	458.5	4.8	591	1 ALU8_HUMAN	P39195 homo sapien
C 8	448	4.7	593	1 ALU6_HUMAN	P39193 homo sapien
9	431.5	4.5	585	1 ALU5_HUMAN	P39192 homo sapien
C 10	426.5	4.4	587	1 ALU2_HUMAN	P39189 homo sapien
C 11	410.5	4.3	587	1 ALU2_HUMAN	P39189 homo sapien
C 12	385	4.0	587	1 ALU3_HUMAN	P39190 homo sapien
C 13	376.5	3.9	585	1 ALU5_HUMAN	P39192 homo sapien
C 14	362.5	3.8	587	1 ALU3_HUMAN	P39190 homo sapien
C 15	310.5	3.2	603	1 ALU4_HUMAN	P39191 homo sapien
C 16	255.5	2.7	603	1 ALU4_HUMAN	P39191 homo sapien
C 17	224	2.3	418	1 YYY1_HUMAN	P49646 homo sapien
C 18	204.5	2.1	418	1 YYY1_HUMAN	P49646 homo sapien

19	139.5	1.4	124	1	YYY3_HUMAN	P20931 homo sapien
C 20	136	1.4	841	1	NEK4_HUMAN	P51957 homo sapien
C 21	129.5	1.3	629	1	2195_HUMAN	O14628 homo sapien
C 22	127	1.3	841	1	NEK4_HUMAN	P51957 homo sapien
C 23	122.5	1.3	881	1	PRP2_HUMAN	Q39959 homo sapien
C 24	113	1.2	619	1	REL_HUMAN	Q04864 homo sapien
C 25	109.5	1.1	501	1	GYG2_MOUSE	O15488 homo sapien
C 26	107	1.1	465	1	SEP3_MOUSE	Q921S5 mus musculus
C 27	105	1.1	1371	1	UBPJ_HUMAN	O94966 homo sapien
C 28	99	1.0	77	1	Q300_MOUSE	Q02722 mus musculus
C 29	99	1.0	369	1	TA2R_HUMAN	P21731 homo sapien
C 30	98	1.0	438	1	SGCE_HUMAN	O43556 homo sapien
C 31	98	1.0	447	1	KBF3_HUMAN	Q04860 homo sapien
C 32	97.5	1.0	590	1	NU5M_TRYBB	P04540 trypanosoma
C 33	96	1.0	153	1	YH17_YEAST	P38898 saccharomyc
C 34	95.5	1.0	1259	1	EH01_HUMAN	P48553 homo sapien
C 35	93	1.0	169	1	YJ85_YEAST	P47152 saccharomyc
C 36	91.5	0.9	491	1	YEAI_ECOLI	P76236 escherichia
C 37	91.5	1.0	533	1	YADC_SCHPO	Q09837 schizosacch
C 38	90	0.9	447	1	KBF3_HUMAN	Q04860 homo sapien
C 39	90	0.9	718	1	RHG8_HUMAN	Q9nsq0 homo sapien
C 40	88	0.9	453	1	VG_DROME	Q26366 drosophila
C 41	88	0.9	741	1	RED1_HUMAN	P78563 homo sapien
C 42	87.5	0.9	319	1	TCB1_RABIT	P06333 oryctolagus
C 43	87	0.9	510	1	YH43_YEAST	Q03218 saccharomyc
C 44	86	0.9	230	1	UREE_YEREN	P42869 versinia en
C 45	86	0.9	494	1	HMBC_DROME	P09081 drosophila

ALIGNMENTS

RESULT 1					
ALU7_HUMAN					
ID	ALU7_HUMAN	STANDARD;	PRT;	593 AA.	
AC	P39194;				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Alu subfamily SQ sequence contamination warning entry.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=95021756; PubMed=7935834;				
RA	Claverie J.-M., Makalowski W.;				
RT	"Alu alert."				
RL	Nature 371:752-752(1994).				
RN	[2]				
RP	CONCEPT.				
RA	MEDLINE=92241891; PubMed=1572661;				
RA	Claverie J.-M.;				
RT	"Identifying coding exons by similarity search: alu-derived and other				
RT	potentially misleading protein sequences.";				
RL	Genomics 12:838-841(1992).				
RN	[3]				
RP	ALU FAMILIES CLASSIFICATION.				
RA	MEDLINE=88333009; PubMed=3138422;				
RA	Quentin Y.;				
RT	"The Alu family developed through successive waves of fixation				
RT	closely connected with primate lineage history.";				
RL	J. Mol. Evol. 27:194-202(1988).				
RN	[4]				
RP	ALU FAMILIES CLASSIFICATION.				
RA	MEDLINE=91178815; PubMed=1706781;				
RA	Jurka J., Milosavljevic A.;				
RT	"Reconstruction and analysis of human Alu genes.";				
RL	J. Mol. Evol. 32:103-121(1991).				
CC	-!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE				
CC	THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING				
CC	CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX				
CC	FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU				

CC REPEATS.
CC -I- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -I- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -I- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS. WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -I- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14573; -: NOT_ANNOTATED_CDS.
KW Hypothetical protein
FT DOMAIN 1 97 FRAME-1.
FT DOMAIN 101 196 FRAME-2.
FT DOMAIN 200 295 FRAME-3.
FT DOMAIN 299 395 FRAME-4.
FT DOMAIN 399 494 FRAME-5.
FT DOMAIN 498 593 FRAME-6.
SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F3A6089F CRC64;

Alignment Scores:
Pred. No.: 2,33e-48 Length: 593
Score: 554.00 Matches: 194
Percent Similarity: 33.13% Conservative: 28
Best Local Similarity: 28.96% Mismatches: 130
Query Match: 5.74% Indels: 320
DB: 1 Gaps: 16

US-09-966-880A-9 (1-5514) x ALU7_HUMAN (1-593)

QY 3568 TGGCTCACCCCTGTATCCACGACTCTGGGAGGCTGAGGTGGTGGTGGCTGAGCCC 3627
Db 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerProGluVal 24

QY 3628 TGGAGTTCAGACCCGCTGAGCAACATGGCAAAACCCCTGTCTCTATA-----ACAAAA 3681
Db 25 ArgSerArgProAlaTrpProThrTrp***AsnProValSerThrLysAsnThrLys 44

QY 3682 ATTAGCGGGCATGTGGCATGTGCTGTGGTCCGACGCTACTAGGGG-GCTGAGGCGAGA 3740
Db 45 IleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAlaGluAlaGly 64

QY 3741 GAATCTTTGGACCCAGGAGGTCAAGCTGCAGTGGAGTCTTCCGCCACTGCACCTCC 3800
Db 65 GluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaProLeuHisSer 84

QY 3801 AGCCTGGGTGACGAGCACGACGCTTGCCTCAAAAAAATAAGAGAAAAATTAATAATAA 3860
Db 85 SerLeuGlyAsnLysSer-GluThrProSerGlnLys----- 96

QY 3861 TGGAAACAACACTACAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCTGATATTTTG 3920
Db 96 ----- 96

QY 3921 GTATTTAACTTTTAAAGTCAGGGTCTGTCCACCTGCACCTACATATTATAAATATCAATTCT 3980
Db 96 ----- 96

QY 3981 CAATGTATATCCACACAAAGACTGGTACGTGAATGTTTCATAGTACCTTTATTTCACAAAC 4040
Db 96 ----- 96

QY 4041 CCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAATAAACAATAATGTCGTA 4100
Db 96 ----- 96

QY 4101 TATCCATGCAATGGAATACCAACCCTGCAGTACAAAGGAAGAAAGCTACTTTGGGGATCAATC 4160
Db 96 ----- 96

QY 4161 CCAAAGTCATGACGCTAAATGAAAGAGTCAGACATGAAGGAGGAGATAATGTATGCCATA 4220
Db 96 ----- 96

QY 4221 CGAAATTTAGAAATGAAAGTAACATTATGTACAGAAAGCAAAATCAGGGCGGCATAG 4280
Db 97 -----Lys*****AlaGlyArgG1 104

QY 4281 AGGCTCACACCTGTAAATCCACGACCTTTGAGAGGCCACCTGGG-GAAGATTGCTAGAACATC 4339
Db 104 yGlySerArgLeu***SerGlnHisPheGlyArgProArgValAspHisLeuArgSe 124

QY 4340 AGGAGTTCAAGACCAACGCTGGGCAACACAGTGAACCTCCATTCACCAAAAAATGGGAAA 4399
Db 124 rGlyValArgAspGlnProGlyGlnHisGlyGluThrPro-SerLeuLeuLysIleGlnL 144

QY 4400 AAAAGAAACAAATCAAGTCAGTGGTGTCTCTGGGGAGGGAAGGACTGCAAGAGGGAAGAA 4459
Db 144 ys-----LeuAlaGlyArg- 148

QY 4460 GCTCTGGTGGGTGAGGGTGGTTCATTCAGTGT-----C 4492
Db 149 -----GlyGlyGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluAsnArgL 167

QY 4493 TGTATCTCCTGCTGTGGTAGCAGTTTGGGCTGTTTACATCCAAAAATATTTCGTAGAATTAT 4552
Db 167 euAsnProGlyGlyGly-----GlyCysSerGluProArg-----SerArgH 181

QY 4553 GCATCTTAATGGTGGGAGCTTTACTGTATGTAATATTACCTCAATGTGAAGAAAAATAA 4612
Db 181 isCysThrProAlaTrpAlaThrArg-----AlaLysLeuA 193

QY 4613 TGTGTAAGAAAAGTTTCAATTTCTTTGCCAGCAACGTTTATTCAAATTTCTGAGCCCTTT 4672
Db 193 rgLeuLysLys*****ProGlyAla----- 202

QY 4673 ACTTGCATAATTTCTCTGCACCTTCTGC-----CCCGTACCATTAGGTGACAGCACTAGCTCCA 4729
Db 203 -----ValAlaHisAlaCysAsnProSerThrLeuGly-----GlyArgG 216

QY 4730 CAATTTGGATAAATGCATTTCTGGAAAGACATAGGGCAAAATCCAGGCATCCTGCTGTCG 4789
Db 216 lyGlyTrpIleThr***GlyGlnGlu-----PheGluThrSerLeuAlaA 231

QY 4790 TTTTCATATCAACACACGCTGT-----ACAGCTTGTGTTGCTGTCTGCA 4831
Db 231 snMetValLysProArgLeuTyr***LysTyrLysAsn***ProGlyValValAlaGlyA 251

QY 4832 GCTGCAATGGGCACTCTTGATTTCTTTTAAGGAAACTTGGGTTACCAGAGATTTTCCACAA 4891
Db 251 laCysAsnProSer-----TyrSerGlyGly***GlyArgArgIleAla***ThrA 268

QY 4892 ATGCTATTCAAAATTAGTGTCTTATGATATCAAGACACTGTGCTAGGAGCCAGAAAACAA 4951


```
QY 3682 ATTAGCCGGCATGGTGGCATGTGCTGTGGTCCAGCTACTAGGGG-GCTGAGGCAGGA 3740
Db 45 IleSerArgAlaTrpIrpArgAlaProValProAlaThrArgGluAlaGluAlaGly 64
QY 3741 GAATCTTTGGAGCCAGGAGTCAAGGCTGCACTGAGCAGTGTCTGGCCACTGCACCTCC 3800
Db 65 GlySerLeuGluProGlyArgSerArgLeuGln***AlaValIleAlaProLeuHisSer 84
QY 3801 AGCCTGGGTGACAGGACCAGACTTCCTCCAAATAAAGAGAAAAATTAAGAAAAATAA 3860
Db 85 SerLeuGlyAspArgAlaArgProCysLeuLysLys*****Ala----- 100
QY 3861 TGAACAACAACTACAAGAGCTGTTGCTCTAGATGAGCTACTTAGTTAGGCTGATATTTG 3920
Db 100 ----- 100
QY 3921 GTATTTAACTTTAAAGTCAGGCTGTGTCACCTGCACCTACATATTATAAATATCAATTCT 3980
Db 101 -----GlyArgGlyGlySer-----ArgLeu***Ser 109
QY 3981 CAATGTATATCCACACAAAGACTGCTACGTGAATGTTTCATAGTACCTTTATTCACAAAAC 4040
Db 110 GlnHisPheGlyArgProArg-----ArgGlu-AspHis-Leu-----SerP 123
QY 4041 CCCAAAGTAGAGACTATCCAAATATATCCATCAACAGTGAACAAATAAACAATAATGCTCTA 4100
Db 123 roGlyValArgaspGlnProGlyGlnHisSerGluThrPro----- 136
QY 4101 TATCCATGCAATGGATATACACCTGCAGTACAAAGAGAAAGACTACTTGGGGATCAATC 4160
Db 137 --SerLeuGlnLysIleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGln- 155
QY 4161 CCAAAGTCATGACGCTAAATGAAGAGTACAGATGAA-----GGAGGA- 4204
Db 156 -----LeuLeuGlyArgLeuArgGlnGluaspArgLeuSerProGlyGlyA 171
QY 4205 -----GATAATGTATGCCATACGAAATCTTAGAAAAATGAAAGTAA 4244
Db 171 rgGlyCysSerGluPro***SerArgHisCysThrProAlaTrpAlaThrGluArgAsp 191
QY 4245 CTATTAGTTACAAAGCAAAATCAGGCGAGGCTAGAGGCTCACACCTGTAATCCACGCA 4304
Db 191 roValSer---LysLys*****ProGlyAlaValAlaHisAlaCysasnProSerT 210
QY 4305 CTTTGAGAGCCACGT-GGGAAGATTGCTAGAACTCAGGAGTTCAGAACCCAGCTGGGCA 4363
Db 210 hrLeuGlyGlyArgGlyGlyArgIleThr***AlaGlnGluPheGluThrSerLeuGlyA 230
QY 4364 ACACAGTGAACCTCCATTCTCCACAAAATGGGAAAAAAGAACCAATCAGTGGTGT 4423
Db 230 snIleValLysProArgLeuTyrlsLysLysLysAsn***ProGlyValValAlaArgA 250
QY 4424 CCTGTGGGAGGGAGGAGCTGCNAAGAGGAGAGAGCTCTGGTGGGTGAGGTGTGA 4483
Db 250 laCysSerProSerTyrlsGlyGly***GlyArgArgIleAla***AlaArgGluValG 270
QY 4484 TTCAGGTTCTGTCTCTGAGCTGGTGTAGCAGTTTG-----GGGNGTT 4525
Db 270 luAlaAlaValSerArgAspArgAlaThrAlaLeuGlnProGlyArgGlnSerGluThrL 290
QY 4526 TACATCCAAAATATTCGTAGAATATTGATCTTAATGGTGGAGCTTTACTGTATGTAA 4585
Db 290 euSerGlnLysLys*****PhePheLeuArgGlnGlyLeuAlaLeuSer----- 307
QY 4586 ATTATACCTCAATGAAGAAAAATAATGTGAAGAAAGTTTCAAT-----TCTC 4636
Db 308 -----ProArgLeuGluCysSerGlyAlaIleThrAlaHisCysSerL 322
QY 4637 TTGCCAGCAAAACGTTTAAATTCCTGAGCCCTTTTACTTCGCAAAATCTCTGCACCTCT 4696
Db 322 euaspLeuProGlySerSerAsp-ProProAlaSerAlaSerArgValAlaGlyThr--- 340
QY 4697 GCCCCGTACCAATTAGGTGACAGCACTAGCTCCCAAAATTTGGATAAATGCAATTTCTGGAA 4756
Db 340 ----- 340
QY 4757 AGACTAGGACAAAATCCAGGCATCACTGTGCTGCTTTTCATATCAACACCGCTGTACAGCTT 4816
Db 341 -----ThrGlyAlaArgHisHis-----AlaArgLeu 349
QY 4817 GTTGTCTGTCTGCTCAGCTGCAATGGGACTTCTTGATTTCTTTAAGGAAACT----- 4867
Db 350 IlePheVal-----PhePheValGluThrGlyPheHis 360
QY 4868 TGGGTTACCAGAGTATTTTCCAAAATGCTATTCAAAATTAGCTTATGATATGCAAGACA 4927
Db 361 TyrValAlaGlnAlaGlyLeuGluLeuLeuGlySerSer----- 373
QY 4928 CTGTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCACTCATTTATGTGGGAACAACAT 4987
Db 374 -----AspProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 390
QY 4988 AGCAAGATATTTAGATCATTTTGCATAGTTTAAAAAGACACAGAGTACAAAATCACACAT 5047
Db 391 ArgAlaArg*****Phe-----Phe***Asp 400
QY 5048 GCAATCAGTATAATCCAA-----ATCATGTAATATGTCCTGTACAAAGACTAGAGGA 5101
Db 401 ArgValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArgLeuThrAlaAla 420
QY 5102 ATAAACACAAAGATCTTAACAGATCATTTGCTATTAGACACTAAGTCTTAATTATTATTATTA 5161
Db 421 SerThrSerArgAlaGlnAlaIleLeuLeuProGlnProGlu***LeuGlyLeu--- 439
QY 5162 GACACTATGATATTTGAGATTTAAAAAATCTTAATAATTTTAAATTTTAGAGCTCTCTCTA 5221
Db 439 ----- 439
QY 5222 TTTTTCATAGTATTCAAGTTTGCAATGATCAAGTATTACTCTCTCTTTTCTTTTCTTTT 5281
Db 440 -----GlnAlaArgAlaThrThrProGly***PheLeuTyrl 451
QY 5282 TTTTCTT----- 5287
Db 452 PheLeu***ArgArgGlyPheThrMetLeuProArgLeuValSerAsnSerTrpAlaGln 471
QY 5287 ----- 5287
Db 472 ValIleLeuProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaPro 491
QY 5288 -----TTTTTTTGAGATGGAGCTTTTGGTCTTTGCTGCCCCATGCTGGAGTGGAGTG 5335
Db 492 Gly*****PhePheGluThrGlySerArgSer-ValAlaGlnAlaGlyValGlnTr 511
QY 5336 GCATGAYCATAGCTCACTGCAACCTCCACCTCCTCGGTTCAGCAAGAGCTGTCGCCCTCAG 5395
Db 511 paArgAspHisGlySerLeuGlnProArgProProGlyLeuLysArgSerSerCysLeuSe 531
QY 5396 CCTCCCGGTGATAGGATGAGTACAGCGCCGCCACACACTCGGCTAATGTTGTATTTT 5455
Db 531 rLeuProSerSerTrpAspTyrlArgArgAlaProProArgProAlaAsnPheCysIlePh 551
QY 5456 TAGTAGAGATGGGCTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAGA 5513
Db 551 eCysArgAspGlyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeuLys 570
RESULT 3
ALU6_HUMAN
ID ALU6_HUMAN STANDARD; PRT; 593 AA.
AC P39193;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SP sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



```
Db 157 euLeuGlyArgLeuArgGlnGlnAsnArg-...LeuAsnProGlyGlyGly-...G 173
QY 4520 GGTGTTTACATCCAAAATATTTCGTAGAAATTATGCATCTTAATGGGTGGAGTTTACTGT 4579
Db 173 lycGlyGluProArg-... 178
QY 4580 ATGTAATATTATACCTCAATGTAAAGAAAAAATAATGTGTAGAAAAAGTTTCAATTCCTTGG 4639
Db 178 ----- 178
QY 4640 CCAGCAACGTTATTCAAAATCCTGAGCCCTTTTACTTCGCAAAATCTCTGCACCTTGTGC 4699
Db 179 -----SerArgHisCysThrP 184
QY 4700 CCGTACCATTAGTGACAGCACTAGCTCCACAAATTTGGATAATGATCTTCGAAAGA 4759
Db 184 roala-----Trp-----AlaI 188
QY 4760 CTAGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACACGCTGTACAGCTTGTG 4819
Db 188 hrArgAlaLysLeuArg-----LeuLysLys*****ProGlyAlaValAla----- 204
QY 4820 TTGCTGTCGACGCTCAATGFGGGACTCTTGATTCTTTAAGGAAACTTGGGTTACCAGA 4879
Db 205 -----HisAlaCysAsn----- 208
QY 4880 GTATTTCACAAATGCTATTCAAAATTAGTCTTATGATGCAAGACACTGCTAGGAG 4939
Db 209 -----ProSerThrLeuGlyG 214
QY 4940 CCAGAAAACAAAGAGGAGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTT 4999
Db 214 ly-----ArgGlyGlyArgIle-----Thr***GlyArgGluPheG 226
QY 5000 AGATCATTTTGACTAGTTTAAAGCAGCAGA----- 5031
Db 226 luthrSerLeuThrAsnMetGluLysProArgLeuTyr***LysTyrLysAsn***ProG 246
QY 5032 --CTACAAATACACATGCAATCAGTATATCAATCAATCATGTAATATATGCTGTGAGA 5089
Db 246 lyValValAlaHisAlaCys-----AsnProSerTyrSerGly-Gly***GlyArg 262
QY 5090 AAGACTAGAGGAATAACACAGAATCTTAACAGTCATTGTCTATTAGACACTAAGTCTAA 5149
Db 263 -----ArgIleAla***ThrArgGluAlaGluValAlaValSerArgAsp----- 277
QY 5150 TTATTATTATAGACACTATGATATTGAGATTTTAAATAATCTTTAATATTTTAAATTT 5209
Db 277 ----- 277
QY 5210 AGAGCTCTCTATTTTCCATAGTATTCAGTTTGCACATGATCAAGTATTACTCTTTCT 5269
Db 278 ArgAlaIle-----AlaLeuGlnProGlyGlnGlnGluArgAsnSerValSer 293
QY 5270 TTTTTTTTTTTTTTTTTTTTTTGTAGATGAGGTTTTTGGTCTGTGTGCCATGCTGGAGT 5329
Db 294 -LysLys*****PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyVa 313
QY 5330 GGAATGCATGACATAGCTACTGCAACCTCCACCTCTCGGTTTCAAGCAAAAGCTGTCG 5389
Db 313 lGlnTrpArgAspLeuGlySerProGlnProProProGlyPheLysArgPheSerCy 333
QY 5390 CCTCAGCTCCCGGGTAGTGGGATTACAGCGCGCCACACACACTCGGCTAATGTTG 5449
Db 333 sLeuSerLeuProSerSerTrpAspTyrArgHisAlaProProArgProAlaAsnPheCy 353
QY 5450 TATTTTGTAGATAGGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCTGACCT 5509
Db 353 sIlePheSerArgAspGlyValSerProCysTrpSerGlyTrpSerArgThrProAspLe 373
QY 5510 CAGA 5513
|||||
```

```
Db 373 uArg 374
RESULT 4
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

DR EMBL; U14574; -; NOT_ANNOTATED_CDS.
 KW Hypothetical protein.
 FT DOMAIN 1 96 FRAME-1.
 FT DOMAIN 100 195 FRAME-2.
 FT DOMAIN 199 294 FRAME-3.
 FT DOMAIN 298 393 FRAME-4.
 FT DOMAIN 397 492 FRAME-5.
 FT DOMAIN 496 591 FRAME-6.
 SQ SEQUENCE 591 AA: 64395 MW: AC8154AD8A6B280 CRC64;

Alignment Scores:

Pred. No.: 4, 14e-45 Length: 591
 Score: 522.50 Matches: 215
 Percent Similarity: 34.36% Conservative: 54
 Best Local Similarity: 27.46% Mismatches: 173
 Query Match: 5.41% Indels: 342
 DB: 1 Gaps: 24

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)

QY 3552 AGACAGCCAGGTGGTGGTCCAGCCTGTAATCCAGCACCTCTGGAGGCTGAGGTGG 3611
 Db 3 ArgArgProGluValAlaHisAlaCysAsnProSerThrLeuGlyGly-Arg----- 20
 QY 3612 TGGATCGTTGAGCCCTGGAGTTCAGACCCGCTGAGCAACAGCGCAAAACCCCTGTTTC 3671
 Db 21 ----SerProGluValArgSerSerArgProAlaTrpProThrTrp***AsnProValSe 39
 QY 3672 TATA-----ACAAAAATTAGCCGGCATGGGCGATGGCTGGTCCAGCTACTAG 3725
 Db 39 rThrLysAsnThrLysIleSerArgAlaTrpIrpArgAlaProValIleProAlaThrAr 59
 QY 3726 GGG-GCTGAGGAGGAGATCTTTGGAGCCAGGAGGTCAAGGCTGCACCTGAGCAGTGCT 3784
 Db 59 gGluAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgGln***AlaGluI 79
 QY 3785 TCGGCCACTGCACTCAGCCTGGTGGTGACAGGACGACCTTGCCTCAAAAAATAGAG 3844
 Db 79 eAlaProLeuHisSerSerLeuGlyAspArgAlaArgLeuArgLeuLysLys----- 96
 QY 3845 AAAAAATTAATAATAATGAAACAACTACAAAGAGCTGTTGCTCTAGATGAGCTACTTAG 3904
 Db 96 ----- 96
 QY 3905 TTAGCTGATATTTTGGTATTAACTTTTAAAGTCAGGCTCTGTCACCTGCACATTA 3964
 Db 96 ----- 96
 QY 3965 TTAAATATCAATTCATGTATATCCACAAAGACTGGTACGTGAATGTTTCATAGTA 4024
 Db 96 ----- 96
 QY 4025 CCTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAGTGAACAAA 4084
 Db 96 ----- 96
 QY 4085 TAACAAAAATGTCTATATCCATGCAATGGAATACCACCTGCAGTACAAAGGAAGC 4144
 Db 96 ----- 96
 QY 4145 TACTTGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCAGACATGAAGGAGGA 4204
 Db 96 ----- 96
 QY 4205 GATAATGTATGCCATACGAAATTCAGAAAAATGAAGTAACATTATAGTTACAGAAAGCAA 4264
 Db 97 -----*****AlaGI 101
 QY 4265 ATCAGGCGCATAGAGGCTCACACCTGTAATCCAGCAGCTTTTGAGAGGCCACGTGGGA 4324

Db 101 yGlyGlyArgAlaArgTIPLeuThrProValIleProAlaLeuTrpGlu-----GI 118
 QY 4325 ACATTGCTAGAACTCAGGAGTTCAGACCCAGCCTGGGCAACAGCAGTGAACCTCCATCTC 4384
 Db 118 uAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrProSerLe 138
 QY 4385 CACAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTCTGTGG----- 4430
 Db 138 uLeuLysIleGlnLysLeuAlaGlyArgGlyAlaArgLeu***SerGlnLeuLeuGI 158
 QY 4431 -----GGAGGGGAAGGACTGCAAGAGGAGGAGAA 4459
 Db 158 yArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGly-CysSerGluProArg 178
 QY 4460 GC----- 4461
 Db 178 erArgHisCysThrProAlaTrpAlaThrGluArgAspSerValSerLysLys***** 198
 QY 4462 -----TCTGTGGGTGAGGTGCTGATTGCTCAG----- 4488
 Db 198 **ArgAlaGluAlaGlyArgGlySerArgLeu***SerGlnHisPheGlyArgLysI 218
 QY 4489 --GTTCTGTATCTGACTGTGTCAGTGTG---GGTGTCTTACATCCAAAATATTCG 4543
 Db 218 leThr**GlyGlnGluPheGluThrSerLeuAlaAsnMetValLysProArgLeuTr* 238
 QY 4544 TAGAATTATGATCTTAATGTTGAGTGTACTGTATGTAATTAATTAATTAATTAATTAAG 4603
 Db 238 **Lys-----TyrLysAsn***ProGlyValVala 248
 QY 4604 AAAAAATAATGTGAAGAAAGTTCATCTCTTCCAGCAACCTTATTCAAATCTCT 4663
 Db 248 laArg---AlaCysAsnProSerTrpSerGlyGly**GlyArgArgIleAla***Thra 267
 QY 4664 GAGCCTTCTACTTCCCA-----AATTCTCTGCTCTCTCCCTGCTACCAATTAG 4711
 Db 267 rgLuuAlaGluValAlaValSerArgAspArgAlaThrAlaLeuGlnProGlyArgGln 287
 QY 4712 GTGACAGCACTAGCTCCCAAAATGGATAAATGCATTTCTGGAAAAGACTAGGACAAA 4771
 Db 287 erGluThrProSerGlnLysLys*****PheLeuArgArg----- 302
 QY 4772 TCCAGCACTACCTGTGCTTTCATATCAACACCGCTGACAGCTGTGTTGCTGCTGCA 4831
 Db 303 -----SerLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleSer--Alah 319
 QY 4832 GCTGCAAT-----GGGACTCT----- 4848
 Db 319 isCysAsnLeuArgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaG 339
 QY 4849 -----TGATTCTTTAAGGAACTGG 4870
 Db 339 lyIleThrGlyAlaArgHisAlaArgLeuIlePheValPheLeu---ValGluThrG 358
 QY 4871 GTTACAGAGTATTTCCCAAAATGCTATTCAA---ATTAGTCTTATGATATCAAGACA 4927
 Db 358 lyPheHisValGlyGlnAlaGlyLeuLeuLeuThrSerGlyAspLeuProL 378
 QY 4928 CTGTGCTAGA----- 4938
 Db 378 ysValLeuGlyLeuGlnAla***AlaThrAlaProGlyLeuArgPro*****PheP 398
 QY 4939 -----GCCAGAAAAACAAGAG 4954
 Db 398 he***AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArg---- 416
 QY 4955 GAGGAGAAATCAGTCTATTATGTGGGAACAACATAGCAAGATATTAGATCATTTTGACTA 5014
 Db 417 -----LeuThrA 419
 QY 5015 GTTAAAAAGCAGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAA 5074
 Db 5015 ----- 5074


```
Db 181 CysThrProAlaThrGlu---ArgAspProValSerLysLys*****Pro 199
    |||||
Qy 5272 AAAGAAGAGTAATACCTGATGTCACAACTGTAATACATGTAAGAAATAGAGAGC 5213
    :|||
Db 200 GlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGly- 215
    :|||
Qy 5212 TCTAAATTTTAAATATTAAGATTTTAAATCTCAAAATATCATAGTGTCTAATAATAA 5153
    :|||
Db 216 GlyArgIleThr***AlaGlnGluPhe- 224
    :|||
Qy 5152 TAATPAGACTTAGTGCTCAATGACAACTGCTTAAAGATTCTGTGTTTATTCCTCTAGT 5093
    :|||
Db 225 -----Glu 225
Qy 5092 CTTTCTACAGGCACATATTACATGATTTGGATTATATCTGATGATGCTGATTTTGTGA 5033
    :|||
Db 226 ThrSerLeuGlyAsnIleValLysProArgLeuTyrLys- 238
    :|||
Qy 5032 CTCTGCTGCTTTTAACTAGTCAAAATGATCTAAATATCTTGCTATGTTGTTCCACAT 4973
    :|||
Db 238 ----- 238
Qy 4972 AATGACTGATTTCTCTCTCTTTGTTTCTGCTCCTAGCACAGTGTCTGCAATCAT 4913
    :|||
Db 238 ----- 238
Qy 4912 AAGCACTAATTTGAATAGCATTTGGAAATACTCTGGTAACCCCAAGTTTCTTTAAGAA 4853
    :|||
Db 239 -----LysTyrLysAsn***ProGlyValValAlaArg 249
    :|||
Qy 4852 ATCAAGAGTCCCAATTCAGCTGCAGACAGCAACAGCTGTACAGCGTGGTTCATATG 4793
    :|||
Db 250 AlaCysSerProSerTyrSerGlyGly***GlyArgArgIleAla***AlaArgGluVal 269
    :|||
Qy 4792 AAGCACAAGTGATGCCTGGATTTGTCCCTAGTCTTTTCCAGAAATGCATTTATPCCAAT 4733
    :|||
Db 270 GluAlaAlaValSerArgAspArgAlaThrAlaLeuGlnPro- 283
    :|||
Qy 4732 TTGTGGAGCTAGTGTGTACCTAATGTACGGGCAGAGTGCAGAGAAATTTGCGAAGT 4673
    :|||
Db 284 -----GlyArgGlnSerGluThrLeuSerGln 292
    :|||
Qy 4672 AAAGGGCTCAGGAATTTGAATAACCTTTGCTGGCAAGAGAAATTTGAACTTTTCTTACACA 4613
    :|||
Db 293 Lys-----Lys***** 296
    :|||
Qy 4612 TTATTTTCTTACATTTAGGTATTAATTTACATACAGTAACCTCCACCCATTTAAGATGC 4553
    :|||
Db 297 ***PhePheLeuArgGlnGlyLeuAlaLeu-----SerProArgLeuGluCys 312
    :|||
Qy 4552 ATAATCTACGAAATATTTTGGATGTAACACCCCAAACTGCTACACAGCAGCATACA 4493
    :|||
Db 313 -----SerGlyAlaIleThrAlaHisCys-----SerLeuAspLeu 324
    :|||
Qy 4492 GAACCTGAATCACCACCTCACCACCCAGAGCTTCTCCCTCTTTGCGAGTCTCCCT 4433
    :|||
Db 325 ProGlySerSerAspProProAlaSerArgValAlaGlyThrThrGlyAlaArg 344
    :|||
Qy 4432 CCCCACGAGACACCACTGATTTGCTTTTCCCATTTTGTGGAGAAATGGAGTT 4373
    :|||
Db 345 HisHisAlaArgLeu-----IlePheValPhePheValGluThrGly-Ph 359
    :|||
Qy 4372 TCACGTGTGTCGCCAGCTGCTTGAACCTCTGAGTCTACGAATCTTCCACAGTGC-GC 4314
    :|||
Db 359 eHisTyrValAlaGlnAlaGlyLeuGluLeuLeuGlySerSerAspProProAlaSerAl 379
    :|||
Qy 4313 CTCTCAAGTGTGGGATACAGGTGTGAGCCTCTATGCTGCCCTGATTCCTTCTGT 4254
    :|||
Db 379 aSerGlnSerAlaGlyIleThrGlyValSerHisArgAla----- 392
    :|||
Qy 4253 AACTATAAGTTACTTTTCTAGAAATTTCTGATGGCATAC----- 4211
    :|||
```

```
Db 393 -----Arg*****PhePhe***AspArgValSerLeuCysArgProGlyTrpSe 410
Qy 4210 ---ATTATCTCCTCTCATGCTGACTCTTTCATTTAGCTCATGACTTTGGGATCA 4155
    :|||
Db 410 rAlaValAlaArgSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeu 430
    :|||
Qy 4154 TCCCAACAGTAGCTTCTCCTTTGTACTGCAGCGGTGGTATTCATTCATGATATAGCAC 4095
    :|||
Db 430 uProGlnProProGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly- 447
    :|||
Qy 4094 ATTTTGTGTTTATTTGTTTCTACTTGTGATGATATTTGGATAGTCTCTACTTTGGGTTTGG 4035
    :|||
Db 448 -----***PheLeuTyrPheLeu***Arg-ArgGlyPheThrMetLeuProArgLeuVal 465
    :|||
Qy 4034 TGAATAAGGTACTATGAACATTCACGTACCAAGTCTTTGTGGGATATACATTGAGAATT 3975
    :|||
Db 465 alSer----- 466
Qy 3974 GATATTTTAAATAAGTAGTGCAGGTGCAGACGCCCTGACTTTTAAAGTTTAAATACCAAAAT 3915
    :|||
Db 467 -----AsnSerTrpAlaGlnValIleLeuPro- 475
Qy 3914 ATCAGCCTAATCACTAGTCTCATCTAGGACAAACAGCTCTTGTAGTGTGTTTCCATTTATT 3855
    :|||
Db 476 -----ProArgProProLysValLeuGlyLeuGlnAla***AlaThrAla-Pro 491
    :|||
Qy 3854 TTTAATTTTCTTCTTATTTTTCAGGCAAGGTCTGGTCTCCCTCACCAGGCTGGAGTG 3795
    :|||
Db 492 -----Gly*****PhePheGluThrGlySerArgSerValAlaGlnAlaGlyVal 509
    :|||
Qy 3794 CAGTGGCCAAAGCAGCTGCTCAGTCAGTCAGCTTGCCTCTCGGCTCCAAAGATTCTCCTGC 3735
    :|||
Db 510 GlnTrpArgAspHisGlySerLeuGlnProArgProProGlyLeuLysArgSerCys 529
    :|||
Qy 3734 CTCAGCCC- CCTAGTAGCTGGGACACAGGACACATGCCACCATGCCCGGCTAATTTTGT 3676
    :|||
Db 530 LeuSerLeuProSerSerTrpAspTyrArgArgAlaProProArgProAlaAsnPheCys 549
    :|||
Qy 3675 TAT-----AGAAACAGGTTTGGCATCTGCTCAGGCTGCTTGAATCCAGGCNC 3622
    :|||
Db 550 IlePheCysArgAspGlyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeu 569
    :|||
Qy 3621 AAGCCATCCACCCACCTCAGCTCCAGAGTGTCTGGATTACAGCGCTGAGCCACCCACAC 3562
    :|||
Db 570 Lys***SerSerArgLeuGlyLeuProLysCysTyrPaspTyrArgArgGluProProArg 589
    :|||
Qy 3561 CTGGCT 3556
    :|||
Db 590 ProAla 591
    :|||
RESULT 6
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily S9 sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
```

potentially misleading protein sequences.";
Genomics 12:838-841(1992).
[3]
ALU FAMILIES CLASSIFICATION.
MEDLINE=88333009; PubMed=3138422;
Quentin Y.;
"The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
J. Mol. Evol. 27:194-202(1988).
[4]
ALU FAMILIES CLASSIFICATION.
MEDLINE=91178815; PubMed=1706781;
Jurka J., Milosavljevic A.;
"Reconstruction and analysis of human Alu genes.";
J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (SPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U14573; -; NOT_ANNOTATED_CDS.

DR Hypothetical protein.
KW DOMAIN 1 97 FRAME-1.
FT DOMAIN 101 196 FRAME-2.
FT DOMAIN 200 295 FRAME-3.
FT DOMAIN 299 395 FRAME-4.
FT DOMAIN 399 494 FRAME-5.
FT DOMAIN 498 593 FRAME-6.
SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.: 1.05e-41 Length: 593
Score: 489.50 Matches: 200
Percent Similarity: 36.28% Conservative: 50
Best Local Similarity: 29.03% Mismatches: 187
Query Match: 5.08% Indels: 254
DB: 1 Gaps: 20

US-09-966-880A-9 (1-5514) x ALU7_HUMAN (1-593)
QY 5512 CTGAGTCCAGGAGTTTGAGACCGAGCTGGCCCAACATGGTGAACCCCATCTCTACTATAAA 5453

122 LeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrProSerLeuLeuLys 141
5452 ATACAAACATTAGCCGAGCTGTGGTGGCGGCTGTAATCCCATCTACCCGGAGGAGCTG 5393
142 IleGlnLysLeuAlaGlyArgGlyGlyArgLeu***SerGlnLeuLeuGlyArgLeu 161
5392 AGCGCAGAGCTTTGCTTGAACCCAGGAGGTGGAGTTGCAGTGAGCTATGTCATGCAT 5333
162 ArgGlnGluAsnArgLeuAsnProGlyGlyGlyCysSerGluProArgSerArgHis 181
5332 TCCATCCAGCATGGGCAACAGCAACAACTCCATCTCAAAAACTCAAAAAAANAAAA 5273
182 CysThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLys*****Pro--- 200
5272 AAAAGAAAGAGTAGTAATCTGATCATCTGTCAACTTGAATCTATGGAATAATAGAAGAGC 5213
200 -----
5212 TCTAAATTTTAAATATTAAGAGATTTTAAATCTCAATATCATAGTGTCTAATAATAA 5153
200 -----
5152 TAATTAGACTTAGTGTCTAATGACAATGACTGTGTAAGATTCTTGTGTTTATTCCTCTAGT 5093
201 -----
5092 CTTTCTACA-----GGCACATATTACATGATTGGATTACTGATTGTC 5048
209 ProSerThrLeuGlyGlyArgGlyGlyTrpIleThr***GlyGlnGluPheGluThrSer 228
5047 ATGTGTGATTTTGTACTCTGCTCTCTTTTAACTAGTCAAAATGATCTAAATATCTTGCT 4988
229 LeuAlaAsnMetVal-----LysProArgLeuTyr***** 239
4987 ATGTTGTTCCACATATGACTGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4928
239 -----
4927 TGTCTTGCATATCATAGCACTAATTTGAATAGCATTTGTGGAAATACTCTGCTAACCCA 4868
240 -----LysTyrLysAsn***Pro 245
4867 AGTTTCTTAAAGAAATCAAGAGTCCCATTCGAGCTGCAGACAGCAACAAAGCTGTAC 4808
246 GlyValValAlaGlyAlaCysAsnProSerTyrSerGlyGly***GlyArgArgIleAla 265
4807 AGCGTGTTCATATGAAGACACAGTGCCTGAGATTTGTCCTCTAGTCTTTTCCAGAA 4748
266 ***ThrArgGluAlaGluValAlaVal----- 274
4747 ATGCATTTATCCAAATTTGTGGAGCTAGTGTCTACCTAATGTCACGGGCGCAAGTGCA 4688
275 -----SerArgAspArg 278
4687 GAGAAATTCGGAAGTAAAGGCTCAGGAATTTGAATACGTTTGTGGCAAGAGAAATTGA 4628
279 AlaThrAlaLeuGlnProGlyGlnGluArgAsnSerVal----- 292
4627 AACTTTTCTACATATTATTTTCTTACATGAGGTATTAATTTTACATACAGTAAGTCC 4568
293 SerLysLys*****PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGly 312
4567 ACCCATTTAAGATGCATAATTTCTACGAATATTTTGGATGTAACACCCCAACTCTCTAC 4508
313 ValGlnTrpArg----- 316
4507 CACAGTCAGGATACAGAACCTGAATCACCACCTCACC-----CACACAGAGCTTCTTCCCT 4451
317 -----AspLeuGlySerLeuGlnProProGlyPheLysArgPheSerCys 333
4450 CTTTGCAGCTCTCC-----CCTCCCCACAGGACACCACTGATTTG 4409

```
Db 334 LeuSerLeuProSerSerTyrArgProProArgProAlaAsn----- 351
Oy 4408 CTTTCTTTTTCCTCCATTTTGTGAGAAATGAGCTTCACTGTCTGCCAGGCTGTCT 4349
Db 352 -----PheCysIlePheSerArgAspGlyValSerProCysTyrProGlyTyrPsr 368
Oy 4348 TGAACCTCTGAGTCTTACGAATCTCCACGT-GGCCTCTCAAAGTCTGGGATTACAGG 4290
Db 369 ArgThrProAspLeuArg***SerThrArgLeuGlyLeuProLysCysTyrAspTyrArg 388
Oy 4289 TGTGAGCT-----CTATGCGCTG 4272
Db 389 ArgGluProProArgProAla*****PhePhe***AspGlyValSerLeuLeu 408
Oy 4271 CCCGATTT-----GCTTCTGCTACTATAAGTACTTTCATTT 4233
Db 409 ProArgLeuGlyCysSerGlyAlaIleSerAlaHisCysAsnLeuArgLeuProGlySer 428
Oy 4232 TCTAGAATTTTCGTATGCATACATTATCTCCTCTTCATCTCTGCTCTTCATTTAGCG 4173
Db 429 Ser----- 429
Oy 4172 TCATGACTTTGGGATTCATCCCAAGTAGCTTCTTCTTGTACTGACGGTGGTATTCC 4113
Db 430 -----AspSerProAlaSer-AlaSerArgVal-----AlaGlyIleTh 442
Oy 4112 ATTCATGGATATAGCACATTTTCTTTTGTCTTCTTCTTGTGTGATGATATTTGGATAGT 4053
Db 442 rGlyAlaArgHisAlaArgLeuIlePheValPheLeuVal----- 456
Oy 4052 CTCTACTTTGGGTTTGTGAATAAAGTACTATGAACATTCACGTACCGTCTTTGTGT 3993
Db 457 -----GluThrGlyPheHisHisValGlyGlnAlaGlyLeu----- 468
Oy 3992 GGATATACATGAGATGATATTTTAAATAGTAGTCAGGTGACAGACCCCTGACTTTA 3933
Db 469 -----GluLeuLeu-----ThrSerGlyAspProPro----- 477
Oy 3932 AAAGTTAAATACCAAAATATACGCTAACTAACTAGCTCATCTAGGACACAGCTCTTTG 3873
Db 478 -----AlaSerAlaSerGlnSerAlaGlyIleTh 487
Oy 3872 TAGTGTGTTCCATTTTAAATTTTCTTCTTATTTTGTAGGCAAGCTCTGGTCT 3813
Db 487 rGly-ValSer---HisArgAlaArg*****PhePheGluThrGluPheArgSerC 506
Oy 3812 GTCACCCAGGCTGAGTGCAGTGCAGCAAGCACTGCTCAGTGCAGCCTTGACCTCCCTGGG 3753
Db 506 ysCysProGlyTyrPsrAlaValAlaArgSerArgLeuThrAlaThrSerAlaSerArgV 526
Oy 3752 CTCCAAGATCTCTGCTCCAGCCCTAGTAG-CTGGGACCACAGGCACATGCCACCA 3694
Db 526 alGlnAlaIleLeuLeuProGlnProProGlu***LeuGlyLeuGlnAlaProAlaThr 546
Oy 3693 TGCCCGGCTAATTTTG-----TTATAGAAGACAGGTTTGGCCATTTGCTCAGGCTGG 3640
Db 546 hrProGly**PheLeuTyrPhe*****ArgArgGlyPheThrMetLeuAlaArgLeuV 566
Oy 3639 TCTTGAACCTCAGGGCTCAAGCGATCCACCACCTCAGCCTCCAGAGTGTGGGATTAC 3580
Db 566 alSerAsnSer***ProGlnValIleHisProProArgProProLysValLeuGlyLeuG 586
Oy 3579 AGCGGTGAGCCACACACCTGGC 3557
Db 586 InAla***AlaThrAlaProGly 593

RESULT 7
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DI 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

Alu subfamily SX sequence contamination warning entry.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=95021758; PubMed=7935834;
Claverie J.-M., Makalowski W.;
"Alu alert."
Nature 371:752-752(1994).
[2]
CONCEPT.
MEDLINE=92241891; PubMed=1572661;
Claverie J.-M.;
"Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
Genomics 12:838-841(1992).
[3]
ALU FAMILIES CLASSIFICATION.
MEDLINE=88333009; PubMed=3138422;
Quentin Y.;
"The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
J. Mol. Evol. 27:194-202(1988).
[4]
ALU FAMILIES CLASSIFICATION.
MEDLINE=91178815; PubMed=1706781;
Jurka J., Milosavljevic A.;
"Reconstruction and analysis of human Alu genes.";
J. Mol. Evol. 32:105-121(1991).
[5]- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
[6]- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
[7]- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
[8]- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
[9]- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U14574; ; NOT_ANNOTATED_CDS.
Hypothetical protein.
DOMAIN 1 96 FRAME-1.
DOMAIN 100 195 FRAME-2.
DOMAIN 199 294 FRAME-3.

FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA: 64395 MW: AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.: 1.66e-38 Length: 591
Score: 458.50 Matches: 191
Percent Similarity: 37.11% Conservative: 68
Best Local Similarity: 27.36% Mismatches: 160
Query Match: 4.76% Indels: 280
DB: 1 Gaps: 22

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)

QY 5512 CTGAGTCTAGGAGTTTGACACGAGCTGGCCACACATGGTGAACCCCATCTCTACTAATAA 5453
DB 121 LeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrProSerLeuLeuLys 140
QY 5452 ATACAAACATAGCCGAGTGTGGTGGCGGCTGTAATCCCATCTACCCGGGAGGCTG 5393
DB 141 IleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyArgLeu 160
QY 5392 AGCGACAGCTTTGTGAACCCAGGAGGTGGAGGTGGAGTGTGAGTGTGATGTCATGCCAT 5333
DB 161 ArgGlnGluAsnArgLeuAsnProGlyGlyGlyCysSerGluProArgSerArgHis 180
QY 5332 TCCATCCAGCATGGCCACACGACCAAACTCCATCTCAAAAAAATAAAAAA 5273
DB 181 CysThrProAlaTrpAlaThr-GluArgAspSerValSerLys*****ArgAl 200
QY 5272 AAAA-----GAAAGAGTAATACTTGATCTGATCTGCAAACTTGAATACATG 5228
DB 200 aGluAlaGlyArgGlySerArgLeu***SerGlnHisPheGlyArg----- 216
QY 5227 GAAATAGAGAGCTCTAAATTTTAAATATTAAGATTTTAAATCTCAATATCAT 5168
DB 217 -LysIleThr***GlyGlnGluPheGluThrSer-----LeuAlaAsnMetVa 232
QY 5167 AGTGTCTAATAATAATAGACTAGTGTCTAATGACAATGACTGTTAAGATTTCTGT 5108
DB 232 I----- 232
QY 5107 GTTATTCTCTAGTCTTTCTACAGGCACATATTTACATGATTGG----- 5062
DB 233 -----LysProArgLeuTyr***LysTyrLysAsn***ProGl 245
QY 5061 -ATTATAGTATGATGCTGATTTTGTACTCTGCTGCTTTT-----TTACTAG 5012
DB 245 yValValAlaArgAlaCysAsnProSerTyrSerGlyGly***GlyArgArgIleAla** 265
QY 5011 TCAAAATGATCTAAATATCTGCTATGTTGTTCCACATATGACTGATTTCTCCTCCTC 4952
DB 265 *ThrArgGluAlaGluValAla-----ValSerArgAspArgAlaThrAl 280
QY 4951 TTTGTTTCTGGTCTCTAGCACAGTGTCTTGCATATCATAAAGCACTAATTGAATAGCAT 4892
DB 280 aLeuGlnProGly----- 284
QY 4891 TTGTGGAAATCTCTGGTAACCCCAAGTTTCTTAAAGAAATCAAGAGTCCCATTCAGC 4832
DB 285 -----ArgGlnSer-GluThrProSerGlnL 293
QY 4831 TGCAGACAGCACACAGCTGTACAGCGTGGTGGATGATGAAGACACAAAGTGTGCTGGA 4772
DB 293 yLys*****PhePheLeuArgArgSer----- 303
QY 4771 TTTTGTCCTAGCTTTTTCAGAAATGCAATTTATCCAAATTTGTGAGCTAGTGTGTCAC 4712
DB 304 --LeuAlaLeu-Ser---ProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsn 321
QY 4711 CTAAATGTCGGGCGAAGTGCAGAGAAATTTGCGAAGTAAGGGCTCAGGAATTTGAAT 4652
DB 111 : : : 111 111

DB 322 LeuArgLeuProGlySerSer----- 328
QY 4651 AACGTTTGTGCGCAAGAGAATTGAAACTTTTCTTACACATATTATTTTCTTACATTGAGG 4592
DB 328 ----- 328
QY 4591 TATAATTTACATACAGTAATAACTCCACCCATTTTAAGATGCATAATTCTACGAATATTTTGT 4532
DB 328 ----- 328
QY 4531 GATGTAAACACCCCAAACTGCTACCACAGTCAGGATACAGAACCTGAATCACCACCTCA 4472
DB 329 -----AspSerProAlaSer 333
QY 4471 CCCCACAGAGCTTTCTCCTCTTTTGCAGTCTTCCCTCCACAGGACAACACTGAT 4412
DB 334 -AlaSerArgValAlaGlyIle-----ThrGlyAlaArgHisH1 346
QY 4411 TTGCTTTCTTTTTCCTTTTTCCTTTTGGAGAAATGGAGTTTCACTGTGTTCCCGAGGCTGG 4352
DB 346 sAlaArgLeuIlePheValPheLeuValGluThrGlyPheHisValGlyGlnAlaG1 366
QY 4351 TCTTGAACCTCTGAGTCTTACCAATCTCCACAGTGGCCTCTCAAAAGTGTCTGGATTACA 4292
DB 366 yLeuGluLeuLeuThrSerGlyAspLeuProPro-----LysValLeuGlyLeuG1 383
QY 4291 GGTGTGAGCCTCTATGCTGCTGCTGCTTCTTCTGTAACTATAAGTTACTTTTCATTTT 4232
DB 383 nAla***AlaThrAlaPro----- 389
QY 4231 CTAGAATTTGCTATGGCATACATATCTCTCTCTCATGTCTGACTCTTTTCATTTACCGT 4172
DB 389 ----- 389
QY 4171 CATGACTTTGGGATTTCATCCCAAGTAGCTTCTTC-----CTTTGTA 4130
DB 390 -----GlyLeu-ArgPro*****PhePhe***AspGlyValSerLeuCysA 406
QY 4129 CTGACAGGTGTATTCATTCATGTCATGATATAGCACATTTTGTATTATTGTTCACTTGTG 4070
DB 406 rgProGlyTrpSerAlaValAla----- 413
QY 4069 ATGGATATTTGGATAGTCTCTACTTTGGGGTTTGTGAATAAGGTACTATGACATTCAC 4010
DB 414 -----ArgSerArgLeuThrAlat 420
QY 4009 CGTACCAGTCTTTGTGTGGATATATATGAGAATTTGATATTTAATAATGTAGTGCAGGT 3950
DB 420 hrSerAlaSer-----ArgValGlnAlaIleLeuLeuProGlnProp 434
QY 3949 GACAGACCTCTGACTTTTAAAGTTAAA----- 3924
DB 434 roGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPhe*** 454
QY 3923 --TACCAAAATATACAGCTTAACCTAGTAGTCTATCTAGGACACACAGCTCTTTGTAGTGT 3866
DB 454 **ArgArgGlyPheThrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIleP 474
QY 3865 TT---CCATTTTATTTTAAATTTT-----TCTT 3842
DB 474 heLeuProLysCysTrpAspTyrArgArgGluProProArgProAlaSerAlaArg*** 494
QY 3841 CTTATTTTGTGGCAAGGTCTGCTGCTGTACCCAGGCTGGAGTGCAGTGGCGGCAAGC 3782
DB 494 ****PhePheGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspL 514
QY 3781 ACTGCTCAGTCGACGCTTGACCTCTGCTGCTCCAAAGATCTCTCCTCAGCCCC-CCTA 3723
DB 514 euGlySerLeuGlnProProProGlyPheLysArgPheSerCysLeuSerLeuProS 534
QY 3722 GTAGCTGGGACACAGCAGCACATGCCACCATGCCCGCTTAATTTTGTGTAT-----AGAA 3669
DB 534 erSerTrpAspTyrArgArgAlaProProArgProAlaAsnPheCysIlePheSerArgA 554


```
QY 3668 ACAGGTTTCCATGCTGCTCAGGCTGCTTGAACCTCCAGGCTCAAGCGATCCACCC 3609
:: ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 554 spGlyValSerProCysTrpProGlyTrpSerArgThrProAspLeuArg***Ser----- 572
QY 3608 ACCTCAGCTCCAGAGTGGTGGATACAGCGCTGAGCCAC 3567
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 573 -----Ser-SerGlnSerAlaGlyIleThrGlyValSerHis 584

RESULT 8
ALU6_HUMAN
ID ALU6_HUMAN STANDARD; PRT; 593 AA.
AC P39193;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SP sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M.; Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=86333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J.; Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -|- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -|- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -|- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -|- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
CC -|- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
```

PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE CODING NUCLEOTIDE SEQUENCE.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14572; -, NOT_ANNOTATED_CDS.

KW Hypothetical protein.

DR DOMAIN 1 97 FRAME-1.

FT DOMAIN 101 196 FRAME-2.

FT DOMAIN 200 295 FRAME-3.

FT DOMAIN 299 395 FRAME-4.

FT DOMAIN 399 494 FRAME-5.

FT DOMAIN 498 593 FRAME-6.

SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:

Pred. No.:	2.01e-37	Length:	593
Score:	448.00	Matches:	196
Percent Similarity:	34.15%	Conservative:	42
Best Local Similarity:	28.12%	Mismatches:	191
Query Match:	4.65%	Indels:	270
Db:	1	Gaps:	21

US-09-966-880A-9 (1-5514) x ALU6_HUMAN (1-593)

QY 5512 CTGAGTCAGGAGTTTGAGACCGAGCTGGCCAAACATGCTGAACCCCATCTCTACTATAAA 5453
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 122 LeuArgSerGlyValArgAspGlnProAspGlnHisGlyGluThrProSerLeuLeuLys 141

QY 5452 ATACAAACATTAGCCGAGTGTGTGGTGGCGCCTGTAAATCCCATCTACCCGGGAGGCTG 5393
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 142 IleGlnLysLeuAlaGlyArgGlyAlaCysLeu***SerGlnLeuLeuGlyArgLeu 161

QY 5392 AGGCACAGCTTTGCTTGAACCCAGGAGGTGGAGTTCAGTGACGTATGTCATGCCAT 5333
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 162 ArgGlnGluAsnArgLeuAsnProGlyGlyGlyCysGlyGluProArgSerArgHis 181

QY 5332 TCCACTCCAGCATGGGCAACAGCAAACTCCCATCTCAAAAAAATAA----- 5285
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 182 CysThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLys*****ProGly 201

QY 5284 -----AAAAAAAAAAAA 5273
:::

Db 202 AlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGlyArgIleThr*** 221

QY 5272 AAAAGAAAGAGTAATACTTGATCATTTGTCAAACTTGTAATATCTATGGAATAAGAGAGC 5213
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 222 GlyArgGluPheGluThr---SerLeuThrAsnMetGlu----- 233

QY 5212 TCTAAATTTTAAATATTAAAGATTTTAAATCTCAAAATCATCATAGTGTCTTAATAATA 5153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 234 -----LysProArgLeuTyr***LysTyrLysAsn----- 243

QY 5152 TAATTAGACTTAGTGTCTTAATGACAATGACTGTTAAGATTTCTGTGTTTATTCCTCTAGT 5093

Db 243 ----- 243

QY 5092 CTTTCTACAGGCACATATTTACATGATTTGGATTATCTACTGATTGATGCTGTGATTTTGTGA 5033
||| |||

Db 244 ----ProGlyValValAlaHis----- 250

QY 5032 CTCGTCTCTTTTAACTAGTCAAAATGATCTAAATATCTTGTCTGTTGTTCCACAT 4973
||| |||

Db 251 ---AlaCys----- 252

QY 4972 AATGACTGATTTCAT 4913

Db 252 ----- 252
 QY 4912 AAGCACTAATTGTAATAGCATTGTGGAAATCTCTGGTAACCAAGTTTCCCTAAAGAA 4853
 Db 253 -----AsnProSerTyrSerGlyGly 259
 QY 4852 ATCAAGAGTCCCATTCAGCTGCAGACAGCAACACAAAGCTGTACAGCGTGGTGTATATG 4793
 Db 260 **GlyArgArgIleAla**ThrArgGluAlaGluValSer---ArgAspArg 278
 QY 4792 AAAGCACAAGTATGCTGCTGATTTTCTCCCTAGTCTTTTCCAGAAATGATTTATCCAAAT 4733
 Db 279 AlaIleAlaLeuGlnPro----- 284
 QY 4732 TTGTGAGCTAGTGTCTCACCTAATGTTACGGGGCAGAGTGCAGAGAAATTTGCAGAGT 4673
 Db 285 -----GlyGlnGlnArgAsnSerValSer 293
 QY 4672 AAAGGCTCAGGAATTTGAATAACGTTTGTGCGCAAGAGAAATTTGAACATTTTCTTACACA 4613
 Db 294 Lys-----Lys***** 297
 QY 4612 TTATTTTCTTACATTGAGGTATAATTTACATACAGTAACACTCCACCCATTTAAAGATGC 4553
 Db 298 ***PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValGlnTrpArg--- 316
 QY 4552 ATAATTCTACGAATATTTTGGATGTAAACACACCCCAACTGCTTACCACAGTCAGGATACA 4493
 Db 317 -----AspLeuGlySer----- 320
 QY 4492 GAACCTGAATACACACCCCTACCCACAGAGCTTCTCCCTCTTTCAGTCTCTCC--- 4436
 Db 321 ---ProGlnProProGlyPheLysArgPheSerCysLeuSerLeuProSerSer 339
 QY 4435 -----CCTCCACAGGACACACCTGATTTGCTTCTTTTCTTCTCC 4394
 Db 340 TrpAspTyrArgHisAlaProProArgProAlaAsn-----PheCys 353
 QY 4393 ATTTTGTGGAGAAATGAGTTTCACTGTGTGCGCCAGGCTGTGAACCTCTGAGTTC 4334
 Db 354 IlePheSerArgAspGlyValSerProCysTrpSerGlyTrpSerArgTrpProAspLeu 373
 QY 4333 TAGCAATCTCCACGCT-GGCCCTCTCAAGTGTGGGATTACAGGTGTGAGCCT----- 4281
 Db 374 Arg***SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArg 393
 QY 4280 -----CTATGCTCGCCTGATTTGCTTTC 4257
 Db 394 ProAla*****PhePhe***AspGlyValSerLeuLeuLeuProArgLeuGlu--- 412
 QY 4256 TGTAACTATAAGTACTTTCTCATTTTCTAGAATTTCTGATGGCATACATTATCTCTCCCTT 4197
 Db 413 CysAsnGlyAlaIleSerAlaHisArgAsnLeu-ArgLeu-----ProGln 427
 QY 4196 CATGCTGACTCTTTCATTTAGCGTATGACTTTGGGATTCATCCCAAGTACTCTCTTC 4137
 Db 427 ySerSerAspSer-----ProAlaSerAlaSerAr 437
 QY 4136 CTTTGTACTGCAGGGTGGTATTCATTGCATGATATAGCACATTTTGTATTATTGTCA 4077
 Db 437 gValAlaGlyIleThrGlyMetArgHisAlaArgLeuIle-PheValPheLeuValG 457
 QY 4076 CTTGTTGATGATATTGGGATCTCTACTTTGGGGTTTGTGAATAAGGTAAGTACTATGA 4017
 Db 457 luthr-----GlyPheLeuHisValGlnAlaG 467
 QY 4016 ACATTACAGTACCAGCTTTTGTGTGATATACATTGAGATTTATTTAAATAGTAG 3957
 Db 467 lLeuGluLeuProThr----- 472
 QY 3956 TGCAGGTGCAGACACCTGACTTTTAAAGTTTAAATACAAATATACAGCTTAAGTAG 3897
 Db 473 ----SerGlyAspProPro-----A 478

QY 3896 CTCACTAGGACAAACAGCTCTTTGTAGTTGTTTCCATTTTATTTTAAATTTTCTTCTAT 3837
 Db 478 laSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*****P 498
 QY 3836 TTTTGTGAGCAAGGTCTGTCTGTCAACCCAGGCTGGAGTGGAGTGGCCCAAGCACTGC 3777
 Db 498 hePheGluThrGluPheArgSerCysCysProGlyTrpSerAlaMetAlaArgSerArgL 518
 QY 3776 TCAGTGCAGCTTGACCTCTCTGGCTCCAAAGATTCTCTGCTCAGCCCTAGTAG-C 3718
 Db 518 eutThrAlaThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProGlu***L 538
 QY 3717 TGGACACACAGGCATGCACCATCCCGGCTAATTTTGT-----TTATAGAAACAGG 3664
 Db 538 euGlyLeuGlnAlaCysAlaThrThrProGly***PheLeuTyrPhe***ArgArgG 558
 QY 3663 GTTTTGCATGTGCTCAGCTGTGTGAACCTCAGGCTCAGCGCATCCACCCACTC 3604
 Db 558 lypheSerMetLeuValArgLeuValSerAsnSerArgProGlnValIleArgProProA 578
 QY 3603 AGCTCCACAGGTGGGATTACAGCGGTGAGCCACACACCTGGC 3557
 Db 578 rgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly 593
 RESULT 9
 ID ALU5_HUMAN STANDARD; PRT; 585 AA.
 AC P39192;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alu subfamily SC sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert."
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences."
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history."
 RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes."
 RL J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES. CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO ACID SEQUENCES.

CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
 CC LOCATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.

CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: U14571; ; NOT_ANNOTATED_CDS.
 CC KW Hypothetical protein.
 CC FT DOMAIN 1 95 FRAME-1.
 CC FT DOMAIN 99 193 FRAME-2.
 CC FT DOMAIN 197 291 FRAME-3.
 CC FT DOMAIN 295 389 FRAME-4.
 CC FT DOMAIN 393 487 FRAME-5.
 CC FT DOMAIN 491 585 FRAME-6.
 CC SQ SEQUENCE 585 AA: 63957 MW: 46E8C4F493650A7 CRC64;

Alignment Scores:

Pred. No.: 1.01e-35 Length: 585
 Score: 431.50 Matches: 187
 Percent Similarity: 34.05% Conservative: 49
 Best Local Similarity: 26.98% Mismatches: 187
 Query Match: 4.47% Indels: 271
 DB: 1 Gaps: 19

US-09-966-880A-9 (1-5514) x ALU5_HUMAN (1-585)

QY 3568 TGGCTCAGCGCTGTAAATCCAGCACTCTGGGAGCTGGAGGTGGGTGGTATCGCTTGCAGCCC 3627
 DB 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerArg-GlyG 24
 QY 3628 TGGAGCTTCAAGCAGCCTGAGCAACATGGCAAAACCCCTGTTCTCTATAC-----AAAA 3681
 DB 24 InGluIleGluThrIleLeuAlaAsnMetValLysProArgLeuTyr***LysTyrLysA 44
 QY 3682 ATTAGCCGGCATGTGGCTGTGCTGTGCTGCCAGCTACTA-GGGGGCTGAGCAGGA 3740
 DB 44 sn**LeuGlyValValAlaArgAlaCysSerProSerTyrSerGlyGly***GlyArgA 64
 QY 3741 GAATCTTTGGAGCCAGGAGTCAAGGCTGCAGCTGAGCAGTGCCTTGGCCACTGCACATCC 3800
 DB 64 rgIleAla***ThrArgGluAlaGluValAlaValSerArgAspArgAlaThrAlaLeuG 84
 QY 3801 AGCCTGGGTGACAGCAGCAGCTTGCCTCAAAAAATAAGA----- 3842
 DB 84 InPro-GlyAspArgAlaArgLeuArgLeuLysLysLys*****AlaGlyArgGlyGly 103
 QY 3843 -----AGAAAAATTAATAATAATGGAACAA 3869
 DB 104 SerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAspHisGluValLysArg 123
 QY 3870 CTACAAAGAGCTGTTGTCTCTAGATGAGCTACTTAGCTAGCTGATATTTTGGTATTTAAC 3929

DB 123 ----- 123
 QY 3930 TTTTAAAGTCAGGGTCTGTCACTGCACCTACATATTATAAATATCAATTTCTCAATGTATA 3989
 DB 124 -----SerArgProSerTrpProThr-Trp***AsnProValSerThrLysAsnThrLy 141
 QY 3990 TCCACACAAAGACTGGTAGCATGTTCATAGTACCTTTATTTCACAAAACCCCAAGTA 4049
 DB 141 sileSerTrpAlaTrpTrp----- 147
 QY 4050 GAGACTATCCAAATATCCATCAACAAGTGAACAATAAACAATAATGTGCTATATCCATCGC 4109
 DB 148 -ArgAlaProValProAlaThrArgGlu-----AlaGluAl 160
 QY 4110 AATGGAATACCACCTGCAGTACAAAGGAAGAGTACTTTGGGATGAATCCCAAGTCA 4169
 DB 160 aGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaProLeuH 180
 QY 4170 TGACGCTAAATGAAGAAGTCAGACATCAAGGAGAGATAATGTATGCCATACGAAATTC 4229
 DB 180 sSerSer-----Le 183
 QY 4230 AGAAATGAAGTAACCTTATAGTTACAGAAACCAATCAGGCAGCATAGAGGCTCACA 4289
 DB 183 uAlaThrGluArgAsp-SerValSerLysLys*****ProGlyAlaValAlaHisA 203
 QY 4290 CCTGTAATCCACAGCACTTTGAGAGCCACGCTGGGAAGATTGCTAGAACCTCAGGAGTCAA 4349
 DB 203 lacysAsnProSerThrLeuGlyGlyArgGlyGlyArgGlyGlyThrArgSerArgAsp-Arg 222
 QY 4350 GACCAGCCTGGGCAACACAGTGAATCTCCATTTCCACAAAAATGGGAAAAAAGAAAGC 4409
 DB 223 AspHisProGlyGlnHisGlyGluThrPro-SerLeuLeuLysIleGlnLys----- 239
 QY 4410 AAATCAGTGGTGTCTCTCTGTTGGGGAGGGAAGGACTGCCAAAGAGGGAAGAGCTCTGGTGG 4469
 DB 240 -----LeuAlaGlyArg-----GlyG 245
 QY 4470 GGTGAGGTGGTGTATTGAGTT-----CTGTATCTCTGA 4502
 DB 245 yAlaArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProG 265
 QY 4503 CTGTGTGAGCAGCTTTGGGTGTATTACATCCAAAAATATTTCGTAGATAATTATGATCTTAA 4562
 DB 265 yGlyGly-----GlyCysSerGluProArg----- 273
 QY 4563 TGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAAATAATGTGTGAAGA 4622
 DB 274 -----SerArgHisCysThrPr 279
 QY 4623 AAGTTTCAATTCCTTTCAGCAAAAGTTTATTCAAATTCCTGAGCCCTTTACTTCGCAAA 4682
 DB 279 oAlaTrpArgGlnSerGluThrProSerGlnLysLys*****PhePheLeuArgAr 299
 QY 4683 TTCTCTGCACCTTCTGCCCGCTACCATTTAGGTGACAGCAGCTAGCTCCACAAATTTGGATAA 4742
 DB 299 gSerLeuAlaLeuSerPro-----GlyTrpSerAlaValAlaArgSerArgLeuTh 316
 QY 4743 TGCATTTCTGGAAGAAGACTAGGACAAATCCAGGCATCCTTGTG----- 4788
 DB 316 rAla-----ThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProProG 333
 QY 4789 ----CTTTTCATATCAACCAAGCTGTACAGCTTGTGTGTCTGCTGCAGCTGCAATGGGA 4844
 DB 333 u***LeuGlyLeuGlnAlaArgAlaThrThr-----Pr 344
 QY 4845 CTCTTGATTTCTTAAAGGAAACTTGGGTTAC-----CAGAGTATTTTCCACAA 4892
 DB 344 oSer***PheLeu-----TyrPhe*****ArgArgGlyPheThrMetLe 359
 QY 4893 TCGTATTCAAATTAGTGTATTGATATGCAAGACACTGTGTAGGAGCCAGAAACAAAG 4952
 DB 359 uAlaArgMetValSerIleSer***ProArgAspProProAlaSerAlaSerGlnSerAl 379

Db 44 sn***ProGlyValValAlaGlyAlaCysSerProSerTyrSerGlyGly***GlyArgA 64
QY 3741 GAATCTTTGGAGCCAGGAGTCAAGCTGACAGTGCAGAGGCTTGGCCCACTGCACCTCC 3800
Db 64 rgMetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaThrAlaLeuG 84
QY 3801 AGCTGGGTGACAGACCAGACCTTGGCTCAAAATAAATAAGAGAAAAATTAATAA 3860
Db 84 lnProGlyArgGlnSerGluThr--ProSerGlnLys-----Lys*****Al 100
QY 3861 TGGAAACACACTACAAAGAGCTGTGTC-----TAGATGAGCTA 3899
Db 100 aGlyArgGlySerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAspHi 120
QY 3900 CTTAGTTAGGTGATATTTTGGTATTTAAAGTCAAGGTCTGTCACCTGACACTA 3959
Db 120 sGluValArg-----ArgSerArgProSerTrpLeuThrAr 132
QY 3960 CATTATTAATAATCAATCTCAATGTATATCCACAAAGACTGGTACGTGAATGTCA 4019
Db 132 g***AsnProValSerThrLysAsnThrLysIleSerArgAlaTrpTtp----- 148
QY 4020 TAGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGA 4079
Db 149 -----ArgAlaProValValProAlaThrArgG 158
QY 4080 ACAAATAACAAAATGTCTATATCCATGCAATGGAATACCACTGCAGTACAAAGGAA 4139
Db 158 u-----AlaGluAlaGlyLupArgGluProGlyArgArgse 171
QY 4140 GAAGCTACTGGGGATGAATCCCAAGTCAATGAGCGTAAATGAAAGAGTACAGATGAAG 4199
Db 171 rLeuGln***AlaGluIleAlaProLeuHisSer----- 183
QY 4200 GAGGAGATAATGTATGCCATGAGAAATCTAGAAAAAGAAAGTAACTATAGTACAGAA 4259
Db 184 -----LeuGlyAspArgAlaArgLeuArgLeuLysL 194
QY 4260 AGCAATCAGGCGAGGATAGAGCTCACACCTGTAATCCAGCAGCTTTGAGAGGCCAG 4319
Db 194 ys*****ProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyArgG 214
QY 4320 TGGGAAGATTTCTAGAACTCAGAGATTCAAGACAGCAGCTGGCCACAGCTGAACTCCA 4379
Db 214 LyGlyArgIleThr-ArgSerGlyAspArgAspHisProGly***HisGlyGluThrPro 233
QY 4380 TTCTCCCAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTGCTGTGGGGAGGGGAA 4439
Db 234 SerLeuLeuLysIleGlnLysLeuAlaGlyArg-----GlyGlyGly 247
QY 4440 GGACTG-----CAAGAGGGAAGAAGCTCTGGTGGGTGAGGTGGTGTTCAGTTCT 4493
Db 248 ArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluAsnGlyValAsnProGlyGly 267
QY 4494 GTATCTGACTGTGTGAGTGGTGGGTGTTTACATCCAAATATTCGTAGAATATG 4553
Db 268 -----GlyAlaCysSerGluProArgSerArgHis----- 277
QY 4554 CATCTTAATGGGTGAGATTACTGTATGTATAATATACCTCAATGTAAAGAAAAAATAAT 4613
Db 278 -----CysThrProAlaTrpAlaThrGluArgAspSer 288
QY 4614 GTGTAGAAAAG---TTTCAATTTCTTGCACAGCAAGCTTATTCAAATTCCTGAGCCCT 4670
Db 289 ValSerLysLys*****PhePheLeuArgArgSerLeu-----AlaLeuSerPro 306
QY 4671 TTACTTCGC-----AAATTTCTGCACTTCTGCCCTTACCATTAGGTGACAGCAC 4721
Db 307 ArgLeuGluCysSerGlyAlaIleSerAla-HisCysLysLeuArgLeuProGlySerAr 326
QY 4722 TAGTCCCAAAATTTGGATAATGCATTTCTGGAAAAAGACTAGGACAAAAATCCAGGCATC 4781
Db 326 gHisSerProAlaSerAlaSerArgValAlaGlyThrThr-Gly-----AlaArgHisH 344

QY 4782 AC-----TTGTGCTTTTCAT-----ATCAACCACG 4805
Db 344 isAlaArgLeuIlePheValPheLeuValGluThrGlyPheHisArgValSerGlnAspG 364
QY 4806 CTGTACAGCTTGTGTGTCTGCTGCAATGGGAGCTCTTGATTTCTTTAAGGAAA 4865
Db 364 lyLeuAspLeuLeuThrSer***SerAlaArgLeuGly-----LeuProLysC 380
QY 4866 CTTGGGTACCAGAGTATTTCCA-----CAATGCTATTCAAATTTAGTGC 4910
Db 380 yStrpAspTyrArgArgGluProProArgProAla*****PhePhe***AspGlyV 400
QY 4911 TTATGATATGCAAGACACTGTGCTAGGAGCAGAAACAAAGAGGAGGAGAAATCAGTCA 4970
Db 400 alSerLeuCysArgPro----- 405
QY 4971 TTATGTGGGAACAACATAGCAAGATATTTAGATCATATTTGACTAGTTAAAAAAGCAGCAG 5030
Db 406 -GlyTrp-----SerAlaValAlaArgSerArgLeuThrAlaSerSerAlaSerA 422
QY 5031 AGTACAAATCACACATCAGTCAATCAGTATATATCCAA-----ATCATGTAAATATGTG 5081
Db 422 rg-----ValHisAlaIleLeuLeuProGlnProGlu***LeuGlyLeuGlnA 439
QY 5082 CTTGTAGAAAAGACTAGAGGAATAACACAGAATCTTAACAGTCATTGTTCATTAGACACT 5141
Db 439 laProAlaThrThrProGly----- 445
QY 5142 AAGTCTAATATTATTATTAGACACTATGATATTTAGATTTTAAATAATCTTTAATATT 5201
Db 446 -----***PheLeuTyrPhe*****ArgArgGlyPheThrValL 459
QY 5202 TAAATTTAGAGCTCTCTCTATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGTATTA 5261
Db 459 eu-----AlaArgMetValSerIleSer***proArgAspProProAlaSerAlaS 476
QY 5262 CTCTTTCTTTTCTTTTCTTTT-----TTTCTTTTCTTTTCTTTTCTTTTCTTTT 5297
Db 476 erGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*****PhePheGluT 496
QY 5298 TGGAGTTTTGTCTTGTTCCTGAGTGGAGTGAATGCATGAYCATGCTACTGCACAA 5357
Db 496 hrGluSerArgSer-ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGln 515
QY 5358 CTCCACCTCTCTGGTTCACAGCAAGCTGTCGCTCAGCCTCCCGGGTAGATGGATPAC 5417
Db 516 AlaProProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAspTyr 535
QY 5418 AGGCGCCACACACACTCGGCTAATCTTTGTTATTATTAGTAGAGATGGGGTTTCACCA 5477
Db 536 ArgArgProProProArgProAlaAsnPheCysIlePheSerArgAspGlyValSerPro 555
QY 5478 TGTGGCCAGCGTGTCTCAAACTCCTGACCTC 5510
Db 556 Cys***ProGlyTrpSerArgSerProAspLeu 566

RESULT 11

ALU2_HUMAN
ID ALU2_HUMAN STANDARD; PRT; 587 AA.
AC P39189;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SB sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;


```
QY 5175 TTGAGATTAAAAAATCTTTAATATTTTAAATTTTAGAGCTCTTCTATTTTTCATAGTA 5234
Db 437 ----- 437
QY 5235 TTCAAGTTTGACAATGATCAAGTATTAATCTTCTTTTCTTTTCTTTTCTTTTCTTTT 5284
Db 438 -----GlnAlaProAlaThrThrProGly***PheLeuTy:Phe*****ArgA 454
QY 5284 ----- 5284
Db 454 rgGlyPheThrValLeuAlaGlyMetValSerIleSer***ProArgAspProProAlas 474
QY 5285 -----TTTTTTT 5291
Db 474 erAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*****PheP 494
QY 5292 TTGAGATGGAGTTTGGTCTTGTGGCCATCTCGATGGATGGATGGATGATGATGATGATG 5351
Db 494 heGluThrGluSerArgSer-ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySer 513
QY 5352 CTGCAACCTCCACCTCTCGGTTCAGCAAGCTGTGCGCTCAGCTCCCGGGTAGATGG 5411
Db 514 LeuGlnAlaProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrp 533
QY 5412 GATTACAGGCCGCCACCACTCGGCTAATGTTTGTATTTTGTAGTAGATGGGTT 5471
Db 534 AspTyArgArgProProLeuArgProAlaAsnPheCysIlePheSerArgAspGlyVal 553
QY 5472 TCACCATGTTGGCAGGCTGTCTCAAACTCTGACCTC 5510
Db 554 SerProPhe***ProGlyTrpSerArgSerProAspLeu 566

RESULT 13
ALU5_HUMAN
ID ALU5_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=911178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
```

FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

-1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

-1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO ACID SEQUENCES.

-1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER, CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS, CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES BEING REPORTED.

-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE CODING NUCLEOTIDE SEQUENCE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14571; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 95 FRAME-1.

FT DOMAIN 99 193 FRAME-2.

FT DOMAIN 197 291 FRAME-3.

FT DOMAIN 295 389 FRAME-4.

FT DOMAIN 393 487 FRAME-5.

FT DOMAIN 491 585 FRAME-6.

SQ SEQUENCE 585 AA; 63957 MW; 46EE8C4F493650A7 CRC64;

Alignment Scores:

Align. No.:	4.78e-30	Length:	585
Score:	376.50	Matches:	176
Percent Similarity:	29.39%	Conservative:	25
Best Local Similarity:	25.73%	Mismatches:	130
Query Match:	3.91%	Indels:	354
DB:	1	Gaps:	13

US-09-966-880A-9 (1-5514) x ALU5_HUMAN (1-585)

QY 5494 GACCAGCTGCCAACATGGTGAACCCCATCTCTACTATAAATACAAACATTACCGAG 5435

Db 223 AsphisProGlyGlnHisGlyGluThrProSerLeuLeuLysIleGlnLysLeuAlaGly 242

QY 5434 TGTGGTGGTGGCGCTGTAATCCCATCTACCGGAGGCTGAGCGACAGCTTTGCTTG 5375

Db 243 ArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlnAsnArgLeu 262

QY 5374 AACCCAGAGGTGGAGTGTGCAGTCAGCTATGTCATGCCATTCACCTCCAGCATGGCA 5315

Db 263 AsnProGlyGlyGlyGlyCysSerGluProArgSerArgHisCysThrProAlaTrp--A 282

QY 5314 ACAAGACCAAACTCCATCTCAAAAAA*****AAAAAAGAGAGTAATACT 5255

Db 282 rgGlnSerGluThrProSerGlnLysLys*****----- 293

QY 5254 TGATCATTTGCAAACTTGAATACTATGAAAAAATAGAAGAGCTCTAAATTTTAAATATT 5195

Db 293 ----- 293

QY	5194	AAAGATTTTAAATCAATATCATAGTGTCTAATAATAATTAGACTTAGTGTCT	5135
Db	294		
QY	5134	AAAGCAATGACTGTAGATTTCTGTGTATTTCTCTAGTCTTTCTACAGGCACATAT	5075
Db	297		
QY	5074	TTACATGATTTGGATTTACTGATTCATGCTGATTTTGTACTCTGCTGCTTTTAAAC	5015
Db	303		
QY	5014	TAGTCAAAATGATCAATATATCTTGCTATGTGTTTCCACATAATGACTGATTTCTCCCTC	4955
Db	304		
QY	4954	CTCTTTGTTTCTGCGTCTCAGACAGTGTCTTTGCATATCATAAAGCACTAAATTTGAATAG	4895
Db	306		
QY	4894	CATTTGTGCAATACTCTGTGAACCAAGTTTCTTAAAGAAATCAAGAGTCCCATTCG	4835
Db	315		
QY	4834	AGCTGCAGACAGCAACAGCTGTACAGCGTGTGATATGAAGCACAAGTGCCT	4775
Db	316		
QY	4774	GGATTTGTCCTAGTCTTTTCCAGAAATGCATTTATCCAAATTTGCGAGCTAGTCTGT	4715
Db	321		
QY	4714	CACCTAATGTCAGGGCAGAGTGCAGAGAAATTTGCGAAGTAAAGGCTCAGGAATTTG	4655
Db	331		
QY	4654	ATAAAGCTTTTCTGCGCAAGAGAAATTTGAACATTTTCTTTACACATTTAT	4607
Db	351		
QY	4606	TTTCTTACATTTGAGGTATATTTACATACAGTAACTCCACCATTTAAGATGCATAT	4547
Db	366		
QY	4546	CTACGAATATTTTGGATGTAAACAC-	4520
Db	378	erAlaGlyIleThrGlyValSerHisArgAlaArg*****PhePhe***AspGlyV	398
QY	4519	CCAACTGTACACAGTCAGGATACAGAACCTGAATCACCACCTCACCACCCAGAGC	4460
Db	398	alSerLeuCysArgGlnAlaGlyValGlnTrp-	410
QY	4459	TTCTTCTCTTTGCACTCTTCCCTCCACAGGACACCACTGATTTGCTTTCTTTT	4400
Db	410	spleuGlySerLeuGln-ProProProGlyPheLysArgPheSerCysLeuSerLeu	429
QY	4399	ProSerSerTrpAspTyrArgArgAlaProProArgProAlaAsnPheCysIlePheSer	449
Db	4384	GAGAATGGAGTTTCACTGTGTTCCCGAGGTGCTTTGAACCTCTGAGTTCTAGCAATCT	4325
QY	4324	TCCACGTGCGCTCTCAAAAGTGTGGATTTACAGGTGTGAGCTCTATGCTGCTCCCTGAT	4265
Db	469	gProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly----	487
QY	4264	TTGCTTTTCTGTAATAAGTTACTTTTCAATTTTCTAGAATTTCTGATGGCATATATC	4205
Db	487	-----	487
QY	4204	TCCTCTTCATGTCTGACTCTTTTCATTTAGCGTCATGACTTTGGGATTTTCATCCCCCAAGTA	4145
Db	487	-----	487
QY	4144	GCTTCTTCTCTTACTGCAGGTGCTATTCCATTCATGCATGATATAGCACATTTTGTGTTA	4085
Db	487	-----	487
QY	4084	TTTGTTCACCTTGTGTGATGATATTTGGATAGTCTCTACTTTGGGTTTGTGTGAATAAGG	4025
Db	487	-----	487
QY	4024	TACTATGAACATTCACGTACCAGTCTTTGTGTGGATATACATTTGAGAATTGATATTTAA	3965
Db	487	-----	487
QY	3964	TAATGTAGTGCAGGTGCAGACCCCTGACTTTTAAAGTTAAATACCAAAATATCAGCCTAA	3905
Db	487	-----	487
QY	3904	CTAAGTAGCTCATCTAGACACAGCTCTTTGTGTAGTTGTTTCCATTTATTTTAATTTT	3845
Db	488	-----	488
QY	3844	CTTCTTATTTTGTAGGCAAGTCTGCTGCTCACCAGGCTGAGTGCAGTGGCGCA	3785
Db	488	*****PhePheGluThrGluSerArgSerValAla-ArgLeuGluCysSerGlyAla	507
QY	3784	AGCACTGCTCAGTGACGCTTGACCTCTGGGCTCCAAAGATTTCTCTGCTCAGGCCCC	3725
Db	508	IleSerAlaHisCysAsnLeuArgLeuProGlySerSerAspSerProAlaSerAlaSer	527
QY	3724	TA-GTAGCTGGGACCACAGGCACATGCCACCATGCCCGGCTAAATTTTGT-ATA	3672
Db	528	ArgValalaclyThrThrGlyAlaArgHisHisAlaGlnLeuIlePheValPheLeuVal	547
QY	3671	GAACAGGTTTTCATGCTGCTCAGCGTGTCTTGAACCTCAGGGCTCAAGCGATCCA	3612
Db	548	GluThrGlyPheHisHisValGlyGlnAspGlyLeuAspLeuLeuThrSer***SerA	567
QY	3611	CCACCTCAGCTCCACAGAGTGTGGATTTACAGGCTGAGCGCTCAGCCACACCTGGCT	3556
Db	567	laargLeuGlyLeuProLysCysTrpAspTyrArgGluProProArgProAla	585
RESULT 14			
ALU3_HUMAN STANDARD: PRT: 587 AA.			
AC	P39150:		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Alu subfamily SBI sequence contamination warning entry.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95021758; PubMed=7935834;		
RA	Claverie J.-M., Makalowski W.;		
RT	"Alu alert."		
RL	Nature 371:752-752(1994).		
RN	[2]		
RP	CONCEPT.		
RX	MEDLINE=92241891; PubMed=1572661;		
RA	Claverie J.-M.;		
RT	"Identifying coding exons by similarity search: alu-derived and other		
RT	potentially misleading protein sequences.";		
RL	Genomics 12:838-841(1992).		
RN	[3]		
RP	ALU FAMILIES CLASSIFICATION.		
RX	MEDLINE=88333009; PubMed=3138422;		
RA	Quentin Y.;		


```

QY 4309 CAAAGTCCTGGATACAGGTGTGACCT----- 4281
Db 378 roLysCysTrpAspTrpArgGluProProArgProAla*****PhePhe***A 398
QY 4280 -----CTATGCTCCGCTGATTTGCTTCTGTAACTATAAGTTTACATTTT 4232
Db 398 spGlyValSerLeuCysArgPro-----GlyTrpSerAlaValAlaGlyS 413
QY 4231 CTAGAATTTCGTGGC-----ATACATTA-----TCTCTCCCT 4197
Db 413 erArgLeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProGlnProProL 433
QY 4196 CATGCTGCTACTCTTCATTTACGTCATGACTTTGGGATTCATCCCAAGTAGCTTCTTC 4137
Db 433 ys***LeuGlyLeuGlnAlaProAlaThrProGly----- 445
QY 4136 CTTTGTACTGCAGGTGGTATTCCATTGCATGATATACACATTTTGGTTATTG--- 4081
Db 446 -----***PheLeuTyrPhe*** 452
QY 4080 -----TTCACCTTTGTGATGATTTGGATAGTCTACTTTGGGTTTGTGA 4032
Db 452 **ArgArgGlyPheThr-----ValL 459
QY 4031 ATAAAGTACTATGAACATTCACGTCACAGCTCTTTGTGGATATACATTGAGAATTGAT 3972
Db 459 euAlaGlyMetValSerIleSer***Pro----- 468
QY 3971 ATTTTAATAATGCTAGTCGAGGTGCAGACGCTGACTTTAAAGTTAAATACCAAAATATC 3912
Db 469 -----ArgAspProPro----- 472
QY 3911 AGCCTAACTAGTACTCATCTAGGACACAGCTCTTTGTAGTCTTTCCATTTATTTT 3852
Db 473 -----AlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis-Arg 487
QY 3851 AATTTTCTCTTATTATTGAGCAAGGTCTGCTGTCACCCAGGCTGGAGTCGAG 3792
Db 488 AlaArg*****PhePheGlyThrGluSerArgSerValAlaGlnAlaGlyValGln 507
QY 3791 TGGCGCAAGCAGCTGCTAGTCAGCAGCTTGACCTCTGGGCTCAAAGATTCTCCCTGCCTC 3732
Db 508 TrpAlaGspLeuGlySerLeuGlnAlaProProGlyPheThrProPheSerCysLeu 527
QY 3731 AGCCC-CTAGTAGTGGGACACAGGACATGCCACCATGCCCGCGTAATTTTGTAT 3673
Db 528 SerLeuProSerSerTrpAspTyrArgArgProProLeuArgProAlaAsnPheCysIle 547
QY 3672 -----AGAAACAGGTTTTTCCATGTTGCTCAGCTGCTTGAACCTCCAGGGCTCAAG 3619
Db 548 PheSerArgAspGlyValSerProPhe***ProGlyTrpSerArgSerProAspLeu--V 567
QY 3618 CGATCCACCCACCTCAGCTCCAGAGTCTGGGATTACAGCGGTGAGCCACACACCTG 3559
Db 567 alIleArgProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProG 587
QY 3558 GC 3557
Db 587 ly 587

RESULT 15
ALU4_HUMAN
ID ALU4_HUMAN STANDARD; PRT; 603 AA.
AC P39191.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SB2 sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

Alignment Scores:

```

RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RX CONCEPT.
RP MEDLINE=92241891; PubMed=1572661;
RX Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RX ALU FAMILIES CLASSIFICATION.
RP MEDLINE=88333009; PubMed=3138422;
RX Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
RP MEDLINE=91178815; PubMed=1706781;
RX Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPTITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LOCATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS. WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14570; -; NOT_ANNOTATED_CDS.
DR DOMAIN 1 98 FRAME-1.
KW Hypothetical protein.
FT DOMAIN 102 199 FRAME-2.
FT DOMAIN 202 300 FRAME-3.
FT DOMAIN 304 401 FRAME-4.
FT DOMAIN 405 502 FRAME-5.
FT DOMAIN 506 603 FRAME-6.
SQ SEQUENCE 603 AA; 65272 MW; B8AAD0AD46BEA114 CRC64;

```

Pred. No.:	3,11e-23	Length:	603
Score:	310.50	Matches:	186
Percent Similarity:	34.74%	Conservative:	52
Best Local Similarity:	27.15%	Mismatches:	206
Query Match:	3.21%	Indels:	246
DB:	1	Gaps:	27
US-09-966-880A-9 (1-5514) x ALU4_HUMAN (1-603)			
QY	3568	TGCTCAGCGCTGTAATCCAGCAGCTCTGGGAGGCTGAGTGGGTGATCGCTTGAGCCC	3627
Db	5	TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGlySer***-GlyG 24	
QY	3628	TGGAGTTCAAGACGAGCTGAGCAACATGGCAACCCCTGTTCTATAA-----CAAA 3680	
Db	24	lnGluIleGluThrIleLeuAlaAsnLysValLysProArgLeuTyr***LysTyrLysL 44	
QY	3681	AATTAGCGGGCGATGGGTGCTGCTGTGGTCCAGCTACTAGGGGG-CTGAGGCGAG 3739	
Db	44	ysLeuAlaGlyArgGlyGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnG 64	
QY	3740	AGATCTTTGGAGCCCGAGGTCAGGCTGCAGCTGAGCAGTCTTGGCCACTGCACCTC 3799	
Db	64	luAsnGlyValAsnProGlySerGlyAlaCysSerGluProArgLeuArgHisCysSerP 84	
QY	3800	CAGC-----CTGGGTGACAGGACGACCTTGCCTCAAAAATAAGAAGAAAAATTA 3852	
Db	84	roGlnSerGlyLeuGlyLysPArgAlaArgLeuArgLeuLysLys*****Ala---- 102	
QY	3853	AAAATAAATGGAACAACACTACAAGAGCTGTGTCTAGATGAGCTACTTAGTGTAGCTG 3912	
Db	102	-----	102
QY	3913	ATATTTTGGTATTAACTTTTAAAGTCAGGCTGTGTCACCTGCAGCTACATATTAAATA 3972	
Db	103	-----GlyArgGlyGlySer-----	ArgL 109
QY	3973	TCAATCTCAATGTATATCCACACAAAGACTGGTA-----CGTAATGTT 4017	
Db	109	eu***SerGlnHisPheGlyArgProArgArgValAspHisGluValArgArgSerArgP 129	
QY	4018	CATAGTACCTTTATTCACAAACCCCAAGTGTAGAGTATCCAAATATCCCATCAACAAGT 4077	
Db	129	roSerTrpLeuThrArg***AsnPro--ValSerThr-----	140
QY	4078	GACAAATAACAAAATGCTATATCCATCCATGGAATACCACCTGCAGTACAAGG 4137	
Db	141	--LysAsnThrLysAsn-----***ProGlyAlaValAlaGlyAlaCysSer----- 155	
QY	4138	AAGAAGTACTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAAGAGTCAGACATG- 4196	
Db	156	--ProSerTyrTrpGlyGly***GlyArgArgMetAla***ThrArgGluAlaGluLeuA 175	
QY	4197	-----AAGGAGGAGATATGTATGCCATACGAAATCTAG-----AAAATGA 4238	
Db	175	laValSerArgAspCysAlaThrAlaValArgSerProAlaTrpAlaThrGluArgAsps 195	
QY	4239	AGTTAACTTATAGTACAGAAACCAATCAGGCGAGCATAGAGGCTCACACCTGTAAATC 4298	
Db	195	erValSer-----LysLys*****ProGlyAlaValAlaHisAlaCysAsnP 212	
QY	4299	CCAGCACTTTGAGAGGCGACGTGGCAAGATGTCTAGAACTCAGGAGTTCAGAGCCAGCCT 4358	
Db	212	roSerThrLeuGlyGlyArgGlyTrpIleMet-ArgSerGlyAspArgAspHisPro 231	
QY	4359	GGGCAACACAGTGAACCTCCATCTCCACAAAATATGGGAATA-----AAAGAAGC 4409	
Db	232	Gly***GlnGlyGluThrProSerLeuLeuLysIleGlnLysIleSerArgAlaArgTrp 251	
QY	4410	AAATCAGTGGTTGCTGTGGGGAGGGAGGACTGCAAGAGGAGAGAGCTCTGGTGG 4469	
Db	252	ArgAlaProValValProAlaThrGlyGluAla---GluAlaGlyGlu-----Trp 267	

QY	4470	GGTGAGGGTGGTGATTGAGTTCTGTATCTGACTGTGTGTAGCAGTGTGGGGTGTATTACA 4529	
Db	268	ArgGluProGlyLysArgSer-----	274
QY	4530	TCCAAAAATATTCGTAGAAATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTA 4589	
Db	275	-----LeuGln***AlaGlu-----IleAlaProLeu 283	
QY	4590	TACCTCAATGTAGAAAAAATAATGTGTAGAAAAATTTCAATTCTCTTGCACGCAACG 4649	
Db	284	GlnSerAlaValArgProGly-----ArgGlnSerGluThrProSerGlnLysLys 300	
QY	4650	TTATTCAAATTCCTGAGCCCTTTTACTTCGCAAAATCTCTGCACCTCTGCCCGACCAATT 4709	
Db	301	*****Phe-----PheLeuArgArgSerLeuAlaLeuSerProArgPro--- 316	
QY	4710	AGGTGACAGCACTAGCTCCACAAATTTGGATAAATGCATTCTCTGGAAGAGACTAGGACAA 4769	
Db	316	-----	316
QY	4770	AATCCAGGCATCACTTGTGCTTTCATATCAACACGCTGTACAGCTTGTGTGTGCTGTCTG 4829	
Db	317	-----AspCysGly-----	319
QY	4830	CAGCTGCAATGGGAGCTCTTGTATTCTTTTAAAGAAA-----	4865
Db	320	--LeuGlnTrpArgAsnLeuGlySerLeuGlnAlaProLeuProGlyPheThrProPhe 338	
QY	4866	-----CTTGGGTACAGAGTATTTCCACAAATGCTATTCCAAATTTAGTGTGTATGATAT 4919	
Db	339	SerCysLeuSerLeuProSerSerTrpAsp-----Tyr 349	
QY	4920	GCAAGACACTGTGTAGGAGCCAGAAA-----ACAAAGAGGAGGAGAAATCAGTCA 4970	
Db	350	ArgArgProProProArgProAlaAsnPheLeuTyrPhe*****ArgArgGlyPheThr 369	
QY	4971	TTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGTACTAGTTAAAAAGCAGCAG 5030	
Db	370	LeuLeuAlaArgMetValSerIleSer***ProHisAspProProAlaSerAlaSerGln 389	
QY	5031	AGTCAAAATCACACATGCAATCAGTATATCCAAATCATGTAATATGTGCCTGTAGAA 5090	
Db	390	SerAla-Gly-----	392
QY	5091	AGACTAGAGGAATAAACACAGAATCTTACACAGTCATTGTTCATTAGACACTAAGCTAAT 5150	
Db	392	-----	392
QY	5151	TATTATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTTAATAATTTTAAAAATTA 5210	
Db	392	-----	392
QY	5211	GAGCTCTCTATTTCATAGTATTCAGTTTGCACATGATCAAGTATTACTCTTTCTTCTT 5270	
Db	393	-----IleThrGlyValSerHisArgAlaArg** 402	
QY	5271	TTTTTTTTTTTTTTTTTTTGTGAGTGGAGTTTGTGCTTGTGTGCCC-----ATGCT 5324	
Db	402	*****PhePhe*****AspGlyValSerLeuCysArgProGlyArgThrAl 419	
QY	5325	GGAGTGAATGGCATGATGATGCTCACTGCAACCTCCACCTCTCGGTGTTCAAGCAAGC 5384	
Db	419	aAspCysSerGlyAlaIleSerAlaHisCysLysLeuArgPheProGlySerArgHis 439	
QY	5385	TGTCGCTTCAGCTCCCGGGTGTAGTGGATTACAGGCCGCCACACACCTCGGCTAAT 5444	
Db	439	rProAlaSerAlaSerProValAlaGlyThrGlyAlaArgHisArgAlaArgLeuI 459	
QY	5445	GTT-TGATTTTTTAGTAGAGATGGGTTCACCATGTTGGCCAGCTGTCTCAACATCC 5503	
Db	459	ePheCysIlePheSerArgAspGlyValSerProCys***ProGlyTrpSerArgSerPr 479	
QY	5504	TGACCTC 5510	

Db |||||||
 479 oAspleu 481

Search completed: June 14, 2003, 18:30:43
Job time : 136.164 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:12:03 ; Search time 338.102 Seconds
(without alignments)
6720.711 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagcagaatacatgtcca.....tcaactoctgacctcagag 5514

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p_model -DEV=xlp
-Q=/cgn2_1/USPTO_spo3/US09966880/runat_14062003_175524_10316/app_query.fasta_1.9493
-DB=SPTRMBL_21 -QMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09966880.ecgn.1.1.978.0runat_14062003_175524_10316 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvivirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
c 1	309.5	3.2	171	4 Q9H728	Q9H728 homo sapien

c 2	293.5	3.0	127	4 Q9H743	Q9h743 homo sapien
c 3	290	3.0	152	4 Q9NX85	Q9nx85 homo sapien
c 4	289	3.0	118	4 Q9H387	Q9h387 homo sapien
c 5	287	3.0	115	6 Q9N083	Q9n083 macaca fasc
c 6	282	2.9	239	4 Q9NX17	Q9nx17 homo sapien
c 7	277	2.9	118	4 Q9H387	Q9h387 homo sapien
c 8	274	2.8	136	4 Q9GNR6	Q9gnr6 homo sapien
c 9	267.5	2.8	152	4 Q9NX85	Q9nx85 homo sapien
c 10	263	2.7	162	4 Q9BYA5	Q9bya5 homo sapien
c 11	260	2.7	72	4 Q8TB48	Q8tb48 homo sapien
c 12	260	2.7	136	4 Q9GNR6	Q9gnr6 homo sapien
c 13	256.5	2.7	163	4 Q9GMM0	Q9gmm0 homo sapien
c 14	255.5	2.7	130	4 Q9HBS7	Q9hbs7 homo sapien
c 15	251	2.6	109	4 Q8WZ39	Q8wz39 homo sapien
c 16	251	2.6	232	4 Q9H5R3	Q9h5r3 homo sapien
c 17	251	2.6	238	4 Q8WT23	Q8wt23 homo sapien
c 18	250.5	2.6	666	4 P78525	P78525 homo sapien
c 19	246	2.5	238	4 Q8WT23	Q8wt23 homo sapien
c 20	246	2.5	239	4 Q9NX17	Q9nx17 homo sapien
c 21	245	2.5	66	4 Q96HL9	Q96hl9 homo sapien
c 22	245	2.5	169	4 Q9H397	Q9h397 homo sapien
c 23	245	2.5	208	4 Q9NW14	Q9nw14 homo sapien
c 24	244.5	2.5	120	4 Q9H6G8	Q9h6g8 homo sapien
c 25	244	2.5	122	6 Q9BGW3	Q9bgw3 macaca fasc
c 26	242.5	2.5	118	4 Q9P195	Q9p195 homo sapien
c 27	241.5	2.5	104	4 Q9P0E3	Q9p0e3 homo sapien
c 28	240.5	2.5	139	4 Q9BVD9	Q9bvd9 homo sapien
c 29	239.5	2.5	123	4 Q9HAD8	Q9had8 homo sapien
c 30	239	2.5	84	4 Q9UHT1	Q9uht1 homo sapien
c 31	239	2.5	151	4 Q9HA67	Q9ha67 homo sapien
c 32	238	2.5	133	4 Q96JR5	Q96jr5 homo sapien
c 33	237	2.5	375	4 Q60448	Q60448 homo sapien
c 34	236	2.4	171	4 Q9H728	Q9h728 homo sapien
c 35	236	2.4	375	4 Q60448	Q60448 homo sapien
c 36	235	2.4	129	4 Q9H9H0	Q9h9h0 homo sapien
c 37	233	2.4	83	4 Q96ID7	Q96id7 homo sapien
c 38	231	2.4	62	4 Q96FS0	Q96fs0 homo sapien
c 39	231	2.4	121	4 Q96N97	Q96n97 homo sapien
c 40	231	2.4	179	4 Q96MD7	Q96md7 homo sapien
c 41	230	2.4	535	4 Q96EB1	Q96eb1 homo sapien
c 42	229.5	2.4	123	4 Q9PIN7	Q9pin7 homo sapien
c 43	229	2.4	666	4 P78525	P78525 homo sapien
c 44	227.5	2.4	139	6 Q9BE58	Q9be58 macaca fasc
c 45	227	2.4	687	4 Q9NXE7	Q9nxe7 homo sapien

ALIGNMENTS

RESULT 1
Q9H728 PRELIMINARY; PRT; 171 AA.
AC Q9H728;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CDNA: FLJ21463 fis, clone COL04765.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Oka T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025116; BAB15071.1; -
SQ SEQUENCE 171 AA; 18436 MW; DDC546D275038FAF CRC64;

Alignment Scores: 1.19e-24 Length: 171
Pred. No.: 171

DT	01-WAR-2001 (TReMBLrel. 16, Created)		
DT	01-NAR-2001 (TReMBLrel. 16, Last sequence update)		
DT	01-WAR-2001 (TReMBLrel. 16, Last annotation update)		
DE	PRO2550.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LIVER;		
RA	Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,		
RA	Xu W., Gao F., Liu M., He F.;		
RT	"Functional prediction of the coding sequences of 75 new genes deduced		
RT	by analysis of cDNA clones from human fetal liver.";		
RL	Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; AF130089; AAG35515.1; --		
SQ	SEQUENCE 118 AA; 13257 MW; 94688870CAC8760D CRC64;		
Alignment Scores:			
Pred. No.:	2,02e-22	Length:	118
Score:	289.00	Matches:	68
Percent Similarity:	71.00%	Conservative:	3
Best Local Similarity:	68.00%	Mismatches:	25
Query Match:	3.00%	Indels:	5
DB:	4	Gaps:	1
US-09-966-880A-9 (1-5514) x Q9H387 (1-118)			
QY	3849 TTTTCTCTTAATTTTTTGAGGCAAGGTCTGGTCCTCACCACCGCTGCAGTG 3790		
Db	5 PhePhePheLeu-PheLeuArgTrpSerPheThrLeuValAlaGlnAlaGlyValGlnTr 24		
QY	3789 GGCGAACACTGCTCAGTCGACGCTTGACCTCCGGCTGCCAAAGATTCTCCGCTCAG 3730		
Db	24 pargAspLeuSerSerProGlnProProProPheArgPheSerCysLeuSe 44		
QY	3729 -CCCCCTAGTAGCTGGGACACGACATGCCACCATGCCCGGCTAAATTTTGT--- 3675		
Db	44 rProProSerSerTrpAspTyArGHisAlaProHisProAlaAsn-PheValPheL 64		
QY	3674 --ATAGAACAAGGTTTGGCATGCTGCTCAGGCTGCTTGAACCTCAGGCGTCAAGCG 3617		
Db	64 euValGlThrGlyPheLeuArgValGlyGlnAlaGlyLeuGluLeuThrSerGlyA 84		
QY	3616 ATCCACCCACTCAGCCCTCCAGAGAGTGCTGGATTACAGCGGTGAGCCACACACC 3561		
Db	84 sProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHistr 102		
RESULT 5			
Q9N083			
ID	Q9N083 PRELIMINARY; . PRT; 115 AA.		
AC	Q9N083		
DT	01-OCT-2000 (TReMBLrel. 15, Created)		
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TReMBLrel. 15, Last annotation update)		
DE	Unamed protein product.		
OS	Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;		
OC	Cercopitheciinae; Macaca.		
OX	NCBI_TaxID=9541;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terso K.,		
RA	Suzuki Y., Sugano S., Hashimoto K.;		
RT	"Isolation of full-length cDNA clones from macaque brain cDNA		
RT	libraries.";		
RL	Submitted (Jul-2000) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; AB046048; BAB01630.1; --		
SQ	SEQUENCE 115 AA; 12702 MW; CB55B257BFBB381A CRC64;		
Alignment Scores:			
Pred. No.:	1.4e-21	Length:	239
Score:	282.00	Matches:	65
Percent Similarity:	67.65%	Conservative:	4
Best Local Similarity:	63.73%	Mismatches:	31
Query Match:	2.93%	Indels:	3
DB:	4	Gaps:	1
US-09-966-880A-9 (1-5514) x Q9NX17 (1-239)			
QY	3865 TTCCATTTTAAATTTTAAATTTTCTTTCTTTTGTGAGCAAGGTCTGTCCTCACCACCGCTG 3806		
Db	109 PheLeuSerPhePhePheLeuPhePhePheLeuArgTrpSerLeuThrLeuSerPro 128		
QY	3805 AGGCTGGAGTGCAGTGGCGGCAAGCAGCTGCTCAGTCAGTCACCTTGCCTCGGCGTCCAAA 3746		
Db	129 ArgLeuGluCySerSerSerAlaIleSerAlaHisCysAlaLeuArgLeuProGlySerSer 148		
Pred. No.:	3.35e-22	Length:	115
Score:	287.00	Matches:	62
Percent Similarity:	70.87%	Conservative:	11
Best Local Similarity:	60.19%	Mismatches:	28
Query Match:	2.98%	Indels:	3
DB:	6	Gaps:	1
US-09-966-880A-9 (1-5514) x Q9N083 (1-115)			
QY	3859 TTATTTTAAATTTTCTTTCTTTTGTGAGCAAGGTCTGTCCTCACCACCGCTG 3800		
Db	1 MetileMetMetIleTyrrLeuPhePheTrpArgGlnGlyLeuAlaLeuSerProArgLeu 20		
QY	3799 GAGTCAGTGGCGCAAGCACTGCTCAGTCAGTCAGCTTGCCTCTCGGCTCCAAAGATTCT 3740		
Db	21 LysCysSerGlyThrIleLeuAlaTyrrCysAsnLeuHisLeuLeuGlySerSerAsnPro 40		
QY	3739 CTGCCTCAGCCCCCTA-GTAGCTGGGACACACGACATGCCACCATGCCCGCTAATT 3681		
Db	41 TyrThrSerAlaSerGlnValAlaGlyThrThrSerAlaCysHisThrTrpLeuIle 60		
QY	3680 TTTGTT-----ATAGAAACAGGTTTGGCATGCTGCTCAGGCTGCTTGAACATCCAG 3627		
Db	61 PheIlePheLeuValGluMetGlyPheTyrrHisValAlaGlnAlaGlyValIysLeuLeu 80		
QY	3626 GGCTCAACGATCCACCCACTCAGCTCCAGCTCCCGAGTGTGGATTACAGCGCTGAGCCAC 3567		
Db	81 SerSerSerAspProProThrLeuAlaPheGlnSerAlaGlyIleThrGlyIleSerHis 100		
QY	3566 CACACCTGG 3558		
Db	101 CystThrTrp 103		
RESULT 6			
Q9NX17			
ID	Q9NX17 PRELIMINARY; . PRT; 239 AA.		
AC	Q9NX17;		
DT	01-OCT-2000 (TReMBLrel. 15, Created)		
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TReMBLrel. 15, Last annotation update)		
DE	CDNA FLJ20489 fis, clone RAU08285.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Watanabe K., Kumagai A., Itakura S., Yamashita M., Tashiro H., Ota T.,		
RA	Suzuki Y., Ohbayashi M., Nishii T., Shibahara T., Tanaka T.,		
RA	Nakamura Y., Isigaki T., Sugano S.;		
RT	"NED0 human cDNA sequencing project.";		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; AK000496; BAA91205.1; --		
SQ	SEQUENCE 239 AA; 25728 MW; 5272FFA2C34214A5 CRC64;		
Alignment Scores:			
Pred. No.:	1.4e-21	Length:	239


```

QY 3745 GATTCTCTGCTCAGCCCCCTA-GTAGCTGGGACCACAGGCACATGCCACCATGCCCGG 3687.
Db :::::::::::::: ||| |||||||| ||||| |||||||| |||||||
149 AnSerProAlaLeuAlaSerGlnValAlaGlyIleThrGlyIleCysHisAlaArg 168
QY 3686 CTAATTTTGT-ATAGAAACAGGTTTGGCATGTGCTCAGGCTGGCTGTGAA 3633
Db |||||||| :::::::::::::: |||||||| |||||||| |||||||| ||||||||
169 GlnIlePheValPheLeuValGluThrGlyPheCysHisValGlyGlnAlaGlyLeuGlu 188
QY 3632 CTCACGGCTCAAGCATCCACCCCTCAGCTCCAGAGTCTCCAGAGTCTGGATTACAGCGTG 3573
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 LeuLeuIleSerGlyAspSerProAlaSerAlaPheGlnSerAlaGlyIleIleGlyVal 208
QY 3572 AGCCAC 3567
Db ||||||
209 SerHis 210

RESULT 7
Q9H387
ID Q9H387 PRELIMINARY; PRT; 118 AA.
AC Q9H387:
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PRO2550.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Xu W., Gao F., Liu M., He F.;
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RT "Functional prediction of the coding sequences of 75 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130089; AAG35515.1; -
SQ SEQUENCE 118 AA; 13257 MW; 94688870CAC8760D CRC64;

Alignment Scores:
Pred. No.: 4.29e-21 Length: 118
Score: 277.00 Matches: 56
Percent Similarity: 74.68% Conservative: 3
Best Local Similarity: 70.89% Mismatches: 19
Query Match: 2.87% Indels: 1
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x Q9H387 (1-118)

QY 5277 TTTTNTTTTNTTTTTCAGATGGAGTTTGGTCTTGTGTCCTGTCCTGAGTGGGAATGG 5336
Db |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
5 PhePhePheLeuPheLeuArgTrpSerPheThrLeuValAlaGlnAlaGlyValGlnTrp 24
QY 5337 CATGAYCATAGTCTACTGCAACCTCCACTCTCTGGTTCAAGCAAGCTGTCGCTCAGC 5396
Db ||| ||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
25 ArgAspLeuSerSerProGlnProProProArgPheLeuArgPheSerCysLeuSer 44
QY 5397 CTCGCGGTAGATGGATTACAGCGCCACACACACTCGGCTTAATGTTGTATTTT 5456
Db ||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
45 ProProSerSerTrpAspTrpArgHisAlaProProHisProAlaAsn-PheValPheLe 64
QY 5457 AGTAGATGGGTTTACCATGTGGCCAGGCTGGTCTCAACTCTGACCTCA 5511
Db |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
64 uValGluThrGlyPheLeuArgValGlyGlnAlaGlyLeuGluLeuThrSer 82

RESULT 8
Q96NR6
ID Q96NR6 PRELIMINARY; PRT; 136 AA.
AC Q96NR6:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CDNA FLJ30278 fis, clone BRACE2002755.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM;
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054840; BAB70813.1; -
SQ SEQUENCE 136 AA; 14980 MW; D2336B649A110163 CRC64;

Alignment Scores:
Pred. No.: 9.5e-21 Length: 136
Score: 274.00 Matches: 54
Percent Similarity: 69.77% Conservative: 6
Best Local Similarity: 62.79% Mismatches: 25
Query Match: 2.85% Indels: 1
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x Q96NR6 (1-136)

QY 5510 GAGTCCAGGAGTTTGAGACCAGCTGGCCAAACATGGTGAACCCCATCTCTACTAAAAAT 5451
Db |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
52 GluValArgSerArgProAlaTrpProThrTrpArgAsnProIleSerThrLysAsn 71
QY 5450 ACAAAACATTAGCCGAGTGTGGTGGCGCTGTAATCCCATCTACCCGGAGGCTGAG 5391
Db ||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
72 ThrLysIleSerGlnThrTrpArgMetProValValLeuAlaThrTrpGluAlaGlu 91
QY 5390 GCGACAGCTTTGCTTGAACCCAGAGGTGGAGTTCGATGTGATGTATGTCCATTC 5331
Db ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 AlaGlyGlu-SerValAsnLeuGlyArgGlyCysSerGluLeuArgArgCysArgCys 111
QY 5330 CACTCCAGCATGGGCAACAGACCAAACTCCATCTCAAAAAAATAAAAAA 5271
Db ||||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
111 sthrProAlaTrpAlaThrArgAlaLysLeuHisLeuArgLysAsnLysAsnLysAsn 131
QY 5270 AAGAAAGAGATAATACT 5255
Db ::||| |||
131 nGlyValThrLysThr 136

RESULT 9
Q9NX85
ID Q9NX85 PRELIMINARY; PRT; 152 AA.
AC Q9NX85:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE KIA0536 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Oka T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000385; BAA91131.1; -
SQ SEQUENCE 152 AA; 16568 MW; 59065F45AAA301B5 CRC64;

Alignment Scores:

```



```
Db      58 SerSerTyrSerProAlaSerAlaSerArgValThrGlyIleThrGlyMetCysHis 77
Qy      3692 GCCGGCTAATTTTGT- ----ATAGAAACAGGGTTTGGCCATGTTGCTCAGGCTGGT 3639
Db      78 AlaGlnLeuIlePheLeuValPheLeuValTysMetGluPheArgHisValGlyGlnThrSer 97
Qy      3638 CTTGAACCTCCAGGCTCAAGCGATCCACCCACCTCAGCTCCAGAGTGGTGGATTACA 3579
Db      98 PheGluLeuLeuAlaSerSerProProAlaSerAlaSerGlnSerAlaGlyIleThr 117
Qy      3578 GCGGTGAGCCACCACACCTGG 3558
Db      118 GlyValSerHisCysAlaTrp 124
```

RESULT 15

```
Q8WZ39 PRELIMINARY; PRT; 109 AA.
AC Q8WZ39;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 12.3 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289553; AAL55737.1; -.
KW Hypothetical protein.
SQ SEQUENCE 109 AA; 12263 MW; 61560F7B74664B55 CRC64;
```

Alignment Scores:

Pred. No.:	3,16e-18	Length:	109
Score:	251.00	Matches:	54
Percent Similarity:	71.08%	Conservative:	5
Best Local Similarity:	65.06%	Mismatches:	23
Query Match:	2.60%	Indels:	1
DB:	4	Gaps:	0

US-09-966-880A-9 (1-5514) x Q8WZ39 (1-109)

```
Qy      5265 TTTCTTTT-----TTTGTGAGATGGAGCTTTTGGTCTTGTGGCCATGCT 5324
Db      27 PheIleLeuValPhePheCysLeuPheLeuArgTrpSer-LeuAlaLeuLeuArgLe 46
Qy      5325 GGAGTGGAAATGGCATGAYCATAGCTACTGCAACCTCCACCTCGCTGGGTTCAAGCAAGC 5384
Db      46 uGluCysSerGlyAlaIleSerAlaHisCysAsnLeuHisLeuLeuGlySerTyrSe 66
Qy      5385 TGTGCGCTCAGCTCCCGGTAGATGGGATTACAGCGCCACCACACCTCGCTAAT 5444
Db      66 rProValSerAlaSerGlnValAlaGlyThrGlyLeuCysHisAlaArgLeuAr 86
Qy      5445 GTTGTATTTTATAGATAGAGATGGGTTTCCACCATGTTGGCAGGCTGCTCAAACTGCT 5504
Db      86 gPheValPheLeuValGlnThrValPheHisHisValGlyGlnAlaGlyLeuLysLeuLe 106
Qy      5505 GACCTCA 5511
Db      106 uAlaSer 108
```

Search completed: June 14, 2003, 18:47:58
Job time : 353.102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:13:03 ; Search time 53.1083 seconds
(without alignments)
6109.704 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagacgaatacatagtcca.....tcaaaactctgacctcagag 5514

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_Spool/US0966880/runat_14062003_175525_10370/app_query.fasta_1.9493
-DB=Issued_Patents_AA -QMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0966880 -CGN_1_1_139 -runat_14062003_175525_10370 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	304	3.2	397	5	PCT-US95-17111A-121
C 2	286	3.0	132	4	US-09-605-785-573
C 3	255	2.6	132	4	US-09-605-785-573
C 4	248.5	2.6	110	4	US-09-227-357-193
C 5	238.5	2.5	99	4	US-09-288-143-168
C 6	237	2.5	375	2	US-08-454-557C-121
C 7	237	2.5	375	2	US-08-340-426D-121
C 8	237	2.5	375	2	US-08-450-673C-121
C 9	236	2.4	375	2	US-08-454-557C-121
C 10	236	2.4	375	2	US-08-340-426D-121
C 11	236	2.4	375	2	US-08-450-673C-121
C 12	221	2.3	397	5	PCT-US95-17111A-121

C 13	221	2.3	1079	3	US-09-058-489-22	Sequence 22, Appl
C 14	212	2.2	1079	3	US-09-058-489-22	Sequence 22, Appl
C 15	198	2.0	56	4	US-09-227-357-577	Sequence 577, App
C 16	192	2.0	99	4	US-09-288-143-168	Sequence 168, App
C 17	187	1.9	90	4	US-09-227-357-171	Sequence 171, App
C 18	181	1.9	110	4	US-09-227-357-193	Sequence 193, App
C 19	170	1.8	122	4	US-09-227-357-285	Sequence 285, App
C 20	169	1.8	122	4	US-09-227-357-285	Sequence 285, App
C 21	166.5	1.7	76	4	US-09-605-785-575	Sequence 575, App
C 22	156.5	1.6	76	4	US-09-605-785-575	Sequence 575, App
C 23	146	1.5	80	4	US-09-227-357-249	Sequence 249, App
C 24	139	1.4	72	4	US-09-227-357-655	Sequence 655, App
C 25	137.5	1.4	52	4	US-09-227-357-537	Sequence 537, App
C 26	136	1.4	388	4	US-09-265-630-11	Sequence 11, Appl
C 27	136	1.4	500	4	US-09-265-630-13	Sequence 13, Appl
C 28	128.5	1.3	80	4	US-09-227-357-249	Sequence 249, App
C 29	127	1.3	416	2	US-08-694-915-4	Sequence 4, Appl
C 30	124	1.3	35	4	US-09-288-143-170	Sequence 170, App
C 31	123.5	1.3	52	4	US-09-227-357-537	Sequence 537, App
C 32	123	1.3	47	4	US-09-227-357-656	Sequence 656, App
C 33	123	1.3	72	4	US-09-227-357-655	Sequence 655, App
C 34	117	1.2	56	4	US-09-227-357-577	Sequence 577, App
C 35	116	1.2	90	4	US-09-227-357-171	Sequence 171, App
C 36	113.5	1.2	70	4	US-09-188-930-131	Sequence 131, App
C 37	113	1.2	619	3	US-09-156-253-48	Sequence 48, Appl
C 38	113	1.2	619	5	PCT-US93-03027-6	Sequence 6, Appl
C 39	111.5	1.2	227	6	5498499-2	Patent No. 5498499
C 40	109	1.1	605	4	US-09-440-936-2	Sequence 2, Appl
C 41	108	1.1	500	4	US-09-265-630-13	Sequence 13, Appl
C 42	104	1.1	579	3	US-08-704-711A-1	Sequence 1, Appl
C 43	104	1.1	579	4	US-09-521-220-1	Sequence 1, Appl
C 44	103.5	1.1	43	4	US-09-461-697-127	Sequence 127, App
C 45	103	1.1	65	4	US-09-227-357-538	Sequence 538, App

ALIGNMENTS

RESULT 1
PCT-US95-17111A-121
; Sequence 121, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:

SEQUENCE CHARACTERISTICS:
 LENGTH: 397 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-17111A-121

Alignment Scores:
 Pred. No.: 1.4e-26
 Score: 304.00
 Matches: 123
 Percent Similarity: 44.51%
 Best Local Similarity: 35.55%
 Query Match: 3.16%
 Indels: 80
 Gaps: 13

US-09-966-880A-9 (1-5514) x PCT-US95-17111A-121 (1-397)

```

QY 4441 CTTCCCTCCACAGGACACCACTGATTTCTTTTCC----- 4394
Db      ||||| ||| :||| ||||| ||||| ||||| |||||
QY 77 ProSerArgProProLysValLeuAspThrGlyLeuSerThrMetProGlyLeuCysLeu 96
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4393 -----ATTTTGTGGAGATGGAGTTTCACTGTGTGCCAGGCTGGTCTTGAACCTCT 4340
Db      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 97 AlaAsnPheCysGlyArgAsnArgValSerLeuMetCysProSerTrpSer-----Pro 114
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4339 GAGTTCTAGCAATCTCCCA-CGTGGCTCTCAAGTGTGGGATTACAGGTGTGAGCCT 4281
Db      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 GluLeuLysGlnSerThrCysLeuSerLeuProLysCysTrpAspTyrArgAla--AlaA 134
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4280 CTATGCTGCCCTGATTTGCTTCTCTAATA-----TAAGTACTTCTTCTTCT 4230
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 134 laValProGlyLeuPheLeuLeuPhePheLeuArgHisArgCysProThrLeuThrGlnA 154
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4229 AGAATTGCTGATG-----CATACATATATCTCTCTCATCTGCTGACTCTTTTCAATTA 4176
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 154 spGluValGlnTrpCysAspHisSerSerLeuGlnProSerThrLeuArgSerSerIle- 173
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4175 GCCTCATGACTTGGGATTCATCCCAAGTAGTCTTCTTGTACTGCAGGTTGGTAT 4116
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 -----LeuLeuProGlnProProLysValAla- 182
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4115 TCATTGCTGATGATAGACATTTTGTGTTTATTTGTTCACTGTTGATGATTTGGAT 4056
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 183 --GlyThrLysAspMetHisHisTyrThrTrpLeu-----PheIlePhe---- 197
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4055 AGTCTCTACTTGGGGTTTGTG----- 4033
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 198 -----IlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyValG 215
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4032 -----AATAAGTACTATGACATTCACATTCAGTACCAGTCTTT----- 3997
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 215 lnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerCysP 235
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3996 -----GTGTGGATAT 3987
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 235 roSerLeuLeuSerSerTrpAspTyrArgProProArgLeuAlaAsnPhePheValP 255
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3986 ACATTGAGAAATGATTTTAAATAGTAGTCAGGTCACAGCCCTGACTTTAAAGTT 3927
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 255 heLeuValGluMetGlyPheHisHisValArgGlnValAspAlaArgSerLeuAspLeuV 275
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3926 AAATACCAAAATATCAGCTAATAAGTAGCTCATCTAGGACACAGCTCTTTGTAGTTG 3867
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 275 alileCysLeuProArgProProLysValLeuGlyLeu---GlnAspValThrProThrA 294
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3866 TTTCCATTTATTTTAAATTTTCTTTATTTTGTAGGCAAGGTCCTGGCTGTCACC 3807
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 294 laArgPro-IlePheAsnPheCys-----LeuPheGluMetGluSerHisSerValThr 311
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3806 CAGGCTGAGTGCAGTGGCGCAAGCAGCTGCTCAGTGTGAGCCTTGACCTCTGGGGTCCAA 3747
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 312 GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLys 331
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 2

US-09-605-785-573
 ; Sequence 573, Application US/09605785
 ; Patent No. 6321716

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Jiang, Yuqi
 APPLICANT: Henderson, Robert A.
 APPLICANT: Kalos, Michael D.
 APPLICANT: Fanger, Gary R.
 APPLICANT: Retter, Marc W.
 APPLICANT: Stolk, John A.
 APPLICANT: Day, Craig H.
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Carter, Darrick
 APPLICANT: Li, Samuel
 APPLICANT: Wang, Aijun
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Hepler, William
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 FILE REFERENCE: 210121.427C16
 CURRENT APPLICATION NUMBER: US/09/605,785
 CURRENT FILING DATE: 2000-06-27
 NUMBER OF SEQ ID NOS: 835
 SOFTWARE: FastSeq for Windows version 3.0
 SEQ ID NO 573
 LENGTH: 132
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-605-785-573

Alignment Scores:
 Pred. No.: 1e-24
 Score: 286.00
 Percent Similarity: 65.42%
 Best Local Similarity: 56.07%
 Query Match: 2.97%
 Indels: 5
 Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-605-785-573 (1-132)

```

QY 3851 AATTTTCTTCTTATTTTGTGCTGCTGACCAAGGCTGCTGACCAAGGCTGGAGTGCAG 3792
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3791 TGGCGCAAGCACTGCTCAGTCAGCTGACCTCCTGGGCTCCAAAGATTCCTCGCCTC 3732
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 45 TrpHisAspLeuSerSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3731 AGCCC-CCTAGTAGCTGGGACACAGGACATGCACCATGCCCGGCTAATTTTGT--- 3676
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
QY 3675 ---TATAGAAACAGGGTTTGGCCATGTTCTGCTCAGGCTGGTCTTGAACTCCAGGCTCAAG 3619
      |||::: ||| |||||||||||||||| |||||||||||:::
Db 85 PheSerArgspgylValSerLeuLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCACCTCAGCTCCAGAGTGTGGGATTACAGCGGTGAGCCACACACCTG 3559
      ::||| ||| ||||||::| ||||| ::| |||||
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTrpArgHisLysProHisPro 124
QY 3558 GCTGTCTGTTCCTCACTTAAAT 3538
      ||| |||||
Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 3
US-09-605-785-573
; Sequence 573, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-573

Alignment Scores:
Pred. No.: 4.61e-21 Length: 132
Score: 255.00 Matches: 50
Percent Similarity: 70.13% Conservative: 4
Best Local Similarity: 64.94% Mismatches: 23
Query Match: 2.64% Indels: 0
DB: Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-605-785-573 (1-132)

QY 5283 TTTTTHTTTGGAGTTTGGTCTGTGTGTCCTGAGTGGAGTGGCATGAY 5342
      ||||||| ||| ||| |||||||::| |||||
Db 28 PhePheLeuArgGlnGlnSerGlyProValAlaGlnAlaGlyValGlnTrpHisAsp 47
QY 5343 CATAGCTCACTGCAACTCCACCTCCTGGGTTTCAAGCAAGCTGTCGCTCAGCCCTCCG 5402
      ||||||| ||| ||| |||||
Db 48 LeuSerSerLeuGlnProLeuProHisArgPheLeuGlnPheSerCysLeuSerLeuPro 67
QY 5403 GGTAGTGGATTACAGCGCCGCCACACCATCTCGGCTAATCTTTGTATTTTAGTAGA 5462
      |||||::| ||| ||||| ||| ||| |||
Db 68 HisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSerPheSerArg 87
QY 5463 GATGGGTTTACCATTGTTGGCCAGGCTGGTCTCAACTCTCTGACCTCAGA 5513
      ||||||| ||| |||||||::| |||||
Db 88 AspGlyValSerLeuLeuCysSerGlyTrpSerLysThrProGlyLeuGln 104
```

```
RESULT 4
US-09-227-357-193
; Sequence 193, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
```



```
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-340-426D-121
Alignment Scores:
Pred. No.: 1,47e-18 Length: 375
Score: 236.00 Matches: 49
Percent Similarity: 65.00% Conservative: 3
Best Local Similarity: 61.25% Mismatches: 27
Query Match: 2.44% Indels: 1
DB: Gaps: 0
US-09-966-880A-9 (1-5514) x US-08-340-426D-121 (1-375)
QY 5276 TTTTGTGAGTGAGATGGTTTTGGTCCTGCCCATCGTCGAGTGAATG 5335
    ||| ||| ||||| ||||| ||| ||| ||||| ||||| :|||
Db 297 PheAsnPheCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGlnTr 316
QY 5336 GCATGAYCATAGTCACTGCACCTCCACTCCCTGGTTCACAAAGCTGCGCTCAG 5395
    | ::: ||||| ||||| ||||| ||||| ||::: |||||
Db 316 pProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLeuSe 336
QY 5396 CCTCCCGGAGTAGGGATTACAGCGCCCCACCACACCTCGGCTTAATGTTGTATTTT 5455
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 336 rLeuProSerSerTrpAspTyrglyHisLeuProHIsProAlaAsnPheCysIlePh 356
QY 5456 TAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACCTCCTGACCTCAGA 5513
    | ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 356 eileAargglyGlyValserProtyrLeuSerGlyTrpSerGlnThrpRoaspLeuArg 375
RESULT 11
US-08-450-673C-121
; Sequence 121, Application US/08450673C
; Patent No. 594888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450.673C
FILING DATE: 30-MAY-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0609.3840004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 121:
SEQUENCE CHARACTERISTICS:
LENGTH: 375 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
```

PCT-US95-17111A-121

Alignment Scores:
Pred. No.: 9, 01e-17 Length: 397
Score: 221.00 Matches: 49
Percent Similarity: 68.75% Conservative: 6
Best Local Similarity: 61.25% Mismatches: 23
Query Match: 2.29% Indels: 2
DB: 5 Gaps: 0

US-09-966-880A-9 (1-5514) x PCT-US95-17111A-121 (1-397)

QY 5276 TTTTCTTTTGTGAGATGGAGTTTGGCTGTGTCGCCATGCTGGAGTGAATG 5335
Db 298 PheAsnPheCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGlnTr 317
QY 5336 GCATGAYCATACCTCACTCCACCTCCACCTCCTGGGTTCAGCAAGCAAGTGTGCGCTCAG 5395
Db 317 pProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLeuSe 337
QY 5396 CCTCCGGGTAGATGGATTACAGCGCCACACACTCGGCTAATGTTTGTATTTT 5455
Db 337 rLeuProSerSerTrpAspTyr-GlyHisLeuHisHisThrProLeuIlePheValPheS 357
QY 5456 TAGTAGAGATGGGGTTTACCATGTTGGCCAGGCTGTGCTCAAACTCCTGACCTCA 5511
Db 357 erLeuGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLysLeuLeuThrSer 375

RESULT 13

US-09-058-489-22

; Sequence 22, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:
Pred. No.: 1, 6e-16 Length: 1079
Score: 221.00 Matches: 53
Percent Similarity: 72.62% Conservative: 8
Best Local Similarity: 63.10% Mismatches: 20
Query Match: 2.30% Indels: 4
DB: 3 Gaps: 1

US-09-966-880A-9 (1-5514) x US-09-058-489-22 (1-1079)

QY 3806 CAGGTGGAGTGCAGTGGCGCAAGCACTGCTCAGTCAGCACCTTGACCTCTCGGCTCCAA 3747
Db 995 ArgAlaGlyMetGlnTrpCysAspLeuSerLeuGlnProProGlyPheLys 1014
QY 3746 AGATTCTCTGCTCAGCCC-CCTAGTAGCTGGGACCACAGCACATGCCACCATGCCCG 3688
Db 1015 ArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArgHisLeuProSerCysPro 1034
QY 3687 GCTAATTTTGTATATA-----GAAACAGGGTTTCCCATGTTGCTCAGCGTGCTTCA 3634
Db 1035 ThrAsnPheCys-IlePheValGluThrGlyPheHisHisValGlyGlnAlaCysLeuGl 1054

QY 3633 ACTCCAGGGCTCAAGCGATCCACCCACCTCAGCCTCCCGAGAGTGTGGGATTACAGCGGT 3574
Db 1054 uLeuLeuThrSerGlyGlyLeuLeuAlaSerAlaSerGlnSerAlaGlyIleThrGlyVa 1074
QY 3573 GAGCCACCAC 3564
Db 1074 lSerHisHis 1077

RESULT 14

US-09-058-489-22
; Sequence 22, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:

Pred. No.: 1, 85e-15 Length: 1079
Score: 212.00 Matches: 44
Percent Similarity: 69.01% Conservative: 5
Best Local Similarity: 61.97% Mismatches: 21
Query Match: 2.19% Indels: 1
DB: 3 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-058-489-22 (1-1079)

QY 5301 AGTTTGGTCTTGTTCGCCATGCTGGAGTGAATGCATGAYCATAGCTCAGTGCACACT 5360
Db 989 SerPheGlnGluSerLeuArgAlaGlyMetGlnTrpCysAspLeuSerLeuGlnPro 1008
QY 5361 CCACCTCCTGGGTTCAGCAAGCTGTCGCTCAGCTCCCGGTAGATGGGATTACAGG 5420
Db 1009 ProProGlyPheLysArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArg 1028
QY 5421 CGCCACACACACACTCGGCTAATGTTGTATTTTAGTAGAGATGGGTTTCACCATGT 5480
Db 1029 HisLeuProSerCysProThrAsnPheCysIlePhe-ValGluThrGlyPheHisHisVa 1048
QY 5481 TGGCCAGGTGGTCTCAAACTCCTGACCTCA 5511
Db 1048 lGlyGlnAlaCysLeuGluLeuLeuThrSer 1058

RESULT 15

US-09-227-357-577
; Sequence 577, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; EARLIER FILING DATE: 1999-01-08
; EARLIER FILING DATE: 1998-07-07
; EARLIER FILING DATE: 1997-07-08
; EARLIER FILING DATE: 1997-07-08

```

; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 577
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-577

```

```

Alignment Scores:
Pred. No.: 1.53e-14 Length: 56
Score: 198.00 Matches: 35
Percent Similarity: 71.70% Conservative: 3
Best Local Similarity: 66.04% Mismatches: 15

```

```

Query Match: 2.05% Indels: 0
DB: 4 Gaps: 0
US-09-966-880A-9 (1-5514) x US-09-227-357-577 (1-56)
QY 5355 CAACCTCCACCTCCTGGGTTCAAGCAAGCTCTCGCCTCAGCCTCCCGGGTAGATGGGAT 5414
   ::::|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 4 GluProProProProArgPheLysArgPheSerCysLeuSerLeuLeuSerSerTrpAsp 23
QY 5415 TACAGGCGCCACACACACTCGGCTAATGTTGTTATTTTGTAGTAGATGGGTTTCA 5474
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 24 TyrArgArgAlaProProHisValAlaIlePheCysThrLeuSerArgAspGlyValLeu 43
QY 5475 CCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGA 5513
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 44 ProHisTrpProGlyTrpSerGlnThrProAspLeuLys 56

```

Search completed: June 14, 2003, 19:00:00
Job time : 91.1083 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:26:00 ; Search time 177.542 Seconds
(without alignments)
5641.044 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagacgaatacatagtgctcc.....tcaaaactcctgacctcagag 5514

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 817286

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ln2p.model -DEV=xlp
-Q=/cgn2_1/USPto_spool/US09966880/runat_14062003_175526_10413/app_query.fasta_1.9493
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09966880_0CGN_1_1_105_@runat_14062003_175526_10413
-NCPu=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA.*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	345.5	3.6	361	10	US-09-995-494-107
2	309	3.2	93	9	US-09-764-891-5337
3	289.5	3.0	107	9	US-09-989-920-233
4	287	3.0	126	9	US-10-082-830-207

c 5	286	3.0	94	9	US-10-001-835-135	Sequence 135, App
c 6	286	3.0	132	9	US-10-012-896-573	Sequence 573, App
c 7	286	3.0	132	9	US-09-895-793-573	Sequence 573, App
c 8	286	3.0	132	9	US-09-895-814-573	Sequence 573, App
c 9	286	3.0	132	10	US-09-759-143-573	Sequence 573, App
c 10	286	3.0	132	10	US-09-780-669-573	Sequence 573, App
c 11	286	3.0	132	10	US-09-822-827-573	Sequence 573, App
c 12	286	3.0	135	9	US-10-012-896-884	Sequence 884, App
c 13	286	3.0	135	9	US-09-895-793-884	Sequence 884, App
c 14	286	3.0	135	9	US-09-895-814-884	Sequence 884, App
c 15	286	3.0	135	10	US-09-759-143-884	Sequence 884, App
c 16	286	3.0	135	10	US-09-780-669-884	Sequence 884, App
c 17	286	3.0	135	10	US-09-822-827-884	Sequence 884, App
c 18	282.5	2.9	84	9	US-10-001-835-160	Sequence 160, App
c 19	282	2.9	239	10	US-09-800-729-193	Sequence 193, App
c 20	282	2.9	310	10	US-09-800-729-190	Sequence 190, App
c 21	279.5	2.9	100	9	US-10-016-157A-171	Sequence 171, App
c 22	279	2.9	109	9	US-09-989-919-105	Sequence 105, App
c 23	278	2.9	144	9	US-10-011-585A-145	Sequence 145, App
c 24	272	2.8	115	10	US-09-995-494-79	Sequence 79, Appl
c 25	270	2.8	183	9	US-09-989-920-245	Sequence 245, App
c 26	267.5	2.8	449	9	US-10-007-280A-140	Sequence 140, App
c 27	265	2.8	118	10	US-09-764-898-157	Sequence 157, App
c 28	264	2.7	107	10	US-10-074-095-510	Sequence 510, App
c 29	264	2.7	107	10	US-09-764-860-510	Sequence 510, App
c 30	263.5	2.7	105	9	US-09-989-919-121	Sequence 121, App
c 31	261	2.7	144	9	US-10-011-585A-145	Sequence 145, App
c 32	258.5	2.7	116	9	US-10-001-835-161	Sequence 161, App
c 33	258.5	2.7	207	9	US-09-866-050A-489	Sequence 489, App
c 34	256.5	2.7	361	10	US-09-995-494-107	Sequence 107, App
c 35	256	2.7	123	9	US-09-764-868-888	Sequence 888, App
c 36	255	2.6	90	9	US-10-001-873-34	Sequence 34, Appl
c 37	255	2.6	132	9	US-10-012-896-573	Sequence 573, App
c 38	255	2.6	132	9	US-09-895-793-573	Sequence 573, App
c 39	255	2.6	132	9	US-09-895-814-573	Sequence 573, App
c 40	255	2.6	132	10	US-09-759-143-573	Sequence 573, App
c 41	255	2.6	132	10	US-09-780-669-573	Sequence 573, App
c 42	255	2.6	132	10	US-09-822-827-573	Sequence 573, App
c 43	255	2.6	135	9	US-10-012-896-884	Sequence 884, App
c 44	255	2.6	135	9	US-09-895-793-884	Sequence 884, App
c 45	255	2.6	135	9	US-09-895-814-884	Sequence 884, App

ALIGNMENTS

RESULT 1
US-09-995-494-107
; Sequence 107, Application US/09995494
; Patent No. US20020127578A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Chen, Sei-yu
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes
; FILE REFERENCE: DEX-0293
; CURRENT APPLICATION NUMBER: US/09/995,494
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/253,176
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 107
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-995-494-107
Alignment Scores:

Pred. No.: 3,74e-25 Length: 361
 Score: 345.50 Matches: 64
 Percent Similarity: 71.88% Conservative: 5
 Best Local Similarity: 66.67% Mismatches: 24
 Query Match: 3.58% Indels: 3
 DB: 10 Gaps: 1

US-09-966-880A-9 (1-5514) x US-09-995-494-107 (1-361)

QY 5235 TTCAAGTTGACAATGATCAAGTATTACTC-----TTTCTTTT-----TTTCTTTT 5285
 Db PheLeuPheAlaAspSerHisLeuGlyLeuAlaPheHisPheAlaPhePhePhePhePhe 285
 QY 5286 PTTTTTTGAGATGGAGTTTGGCTGTGTGCCCATGCTGGAGTGGAAATGGCATGAYCAT 5345
 Db PhePhePheAlaValAlaSerHisProIleAlaGlnAlaGlyValGlnTrpArgAspLeu 305
 QY 5346 ACCTCAGTCACACCTCCACCTCCTGGTTTCAAGCAAAGCTGTGGCTCAGCTCCGGGT 5405
 Db GlySerLeuGlnProProProGlyPheLeuGlnPheLeuCysLeuSerLeuProGly 325
 QY 5406 AGATGGATTACAGCGCCACACACACTCGGCTAATGTTGTATTTTGTAGTAGAT 5465
 Db SerTrpAspTyrArgArgAlaProProArgGlnAlaAsnPheCysIlePheSerArgAsp 345
 QY 5466 GGGGTTTCAACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAGA 5513
 Db GlyValSerProCysTrpThrGlyTrpSerGlnThrProAspLeuArg 361

RESULT 2

US-09-764-891-5337
 ; Sequence 5337, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5337
 ; LENGTH: 93
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; CURRENT APPLICATION NUMBER: US/10/082,830

Alignment Scores:
 Pred. No.: 9,29e-22 Length: 93
 Score: 309.00 Matches: 57
 Percent Similarity: 76.25% Conservative: 4
 Best Local Similarity: 71.25% Mismatches: 18
 Query Match: 3.20% Indels: 1
 DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-891-5337 (1-93)

QY 5273 TTTTTTTTTTTTTTTTTTTTGGATGGAGTTTGGTCTGTGCCCCATGCTGGAGTGA 5332
 Db PhePhePhePhePhePheGluThrGluSerHisSer-ValThrGlnAlaGlyIleG 23
 QY 5333 ATGGCATGAYCATGCTCAGTCAACCTCCACCTCGTGGTTCAGCAAAGCTGTGCGCT 5392
 Db nTrpHisTyrLeuSerSerLeuGlnProProProArgPheLeuPheSerCysLe 43
 QY 5393 CAGCTCCCGGTAGATGGGATTACAGCGCCACACACACCTCGGCTAATGTTGTAT 5452
 Db uSerLeuLeuSerSerTrpAspTyrArgHisThrProHisAlaAlaAsnPheCysI 63
 QY 5453 TTTTAGTAGAGTGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCT 5510
 Db ePheSerArgAspGlyValSerProCysTrpSerGlyTrpSerArgThrProAspLeu 82

RESULT 3

US-09-989-920-233
 ; Sequence 233, Application US/09989920
 ; Patent No. US20020172957A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Macina, Roberto
 ; APPLICANT: Recipon, Herve
 ; APPLICANT: Chen, Sei-Yu
 ; APPLICANT: Sun, Yongming
 ; APPLICANT: Liu, Chenghua
 ; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
 ; FILE REFERENCE: DEX-0291
 ; CURRENT APPLICATION NUMBER: US/09/989,920
 ; CURRENT FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: 60/252,500
 ; PRIOR FILING DATE: 2000-11-22
 ; NUMBER OF SEQ ID NOS: 284
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 233
 ; LENGTH: 107
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 ; US-09-989-920-233

Alignment Scores:
 Pred. No.: 8,49e-20 Length: 107
 Score: 289.50 Matches: 64
 Percent Similarity: 70.59% Conservative: 8
 Best Local Similarity: 62.75% Mismatches: 27
 Query Match: 3.01% Indels: 4
 DB: 9 Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-989-920-233 (1-107)

QY 3856 TTTTAAATTTTCTTCTTATTTTGTAGGCAAGTCTGTGCTCTCACCCAGGCTGGAG 3797
 Db 1 PhePhePhePheLeuLeuPheCys---AspSerLeuAlaLeuSerProArgLeuGln 19
 QY 3796 TGCAGTGGCGCAAGCACCTGCTCAGTCAGCGCTTACCTTCCCTGGGCTCCAAAGATTCTCCT 3737
 Db CysSerGlyThrIleSerAlaHisCysAsnLeuCysLeuLeuGlySerSerAsnSerPro 39
 QY 3736 GCCTCAGCCCCCTA-GTAGCTGGGACCACAGGCACATGCCACCATGCCCGGCTA----- 3684
 Db ValSerAlaSerTrpValAlaGlyThrThrGlyAlaCysHisAlaTrpLeuThrPhe 59
 QY 3683 ATTTTGTATAGAAACAGGGTTTTCCTCATGCTCAGGCTGGTCTTGAACCTCAGGGC 3624
 Db ValPheLeuValGluThrGlyPheHisValGlyGlnAlaGlyLeuGluPheLeuThr 79
 QY 3623 TCAAGCGATCCACCCACCTCAGCCTCCAGAGTGTGGATTACAGGCTGAGCCACCAC 3564
 Db SerGlyAspProProAlaLeuAlaSerGlnSerAlaGluIleThrGlyValSerHisArg 99
 QY 3563 ACCTGG 3558
 Db 100 AlaTrp 101

RESULT 4

US-10-082-830-207
 ; Sequence 207, Application US/10082830
 ; Publication No. US20030077604A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sun, Yongming
 ; APPLICANT: Recipon, Herve
 ; APPLICANT: Salceda, Susana
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Turner, Leah
 ; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
 ; FILE REFERENCE: DEX-0249
 ; CURRENT APPLICATION NUMBER: US/10/082,830

Percent similarity: 99.42%

Conservative: 10

Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: 9 Gaps: 2

US-09-966-880A-9 (1-5514) x US-10-012-896-573 (1-132)

```
QY 3851 AATTTTCTCTTATTTTGTAGGCAAGTCGTGCTCACCAGGCTGGAGTGCAG 3792
      ||||| :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlyValGln 44
      ||| TGCGCAAGCACTGCTCAGTCAGTCAGCTTGACCTCTCGGGCTCCAAAGATTCTCCTCCCTC 3732
      ||| ||| ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3731 AGCCC--CCTAGTAGCTGGACACACAGGCATGCCACCATGCCCGCTAATTTTGT--- 3676
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
      ||| ---TATAGAAACAGGTTTGGCATGTTGCTCAGGCTGCTTGAACCTCCAGGCTCAAG 3619
      ||| :||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3618 CGATCCACCACCTCAGCTCCAGAGTCTGGGATTACAGGCGTCAGCCACACACCTG 3559
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3558 GCTGCTGTTTCACTTTAAAT 3538
      ||| |||||
Db 125 AlaCysHisIleLeuLeuAsn 131
```

RESULT 7

US-09-895-793-573

; Sequence 573, Application US/09895793
; Publication No. US20020192763A1

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.534C2

; CURRENT APPLICATION NUMBER: US/09/895,793

; CURRENT FILING DATE: 2001-06-29

; NUMBER OF SEQ ID NOS: 982

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 573

; LENGTH: 132

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-895-793-573

Alignment Scores:

Pred. No.: 2,05e-19

Score: 286.00

Length: 132

Matches: 60

Percent Similarity: 65.42% Conservative: 10
Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: 9 Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-895-793-573 (1-132)

```
QY 3851 AATTTTCTCTTATTTTGTAGGCAAGTCGTGCTCACCAGGCTGGAGTGCAG 3792
      ||||| :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlyValGln 44
      ||| TGCGCAAGCACTGCTCAGTCAGTCAGCTTGACCTCTCGGGCTCCAAAGATTCTCCTCCCTC 3732
      ||| ||| ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3731 AGCCC--CCTAGTAGCTGGACACACAGGCATGCCACCATGCCCGCTAATTTTGT--- 3676
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
      ||| ---TATAGAAACAGGTTTGGCATGTTGCTCAGGCTGCTTGAACCTCCAGGCTCAAG 3619
      ||| :||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3618 CGATCCACCACCTCAGCTCCAGAGTCTGGGATTACAGGCGTCAGCCACACACCTG 3559
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3558 GCTGCTGTTTCACTTTAAAT 3538
      ||| |||||
Db 125 AlaCysHisIleLeuLeuAsn 131
```

RESULT 8

US-09-895-814-573

; Sequence 573, Application US/09895814

; Publication No. US20020193296A1

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427C26

; CURRENT APPLICATION NUMBER: US/09/895,814

; CURRENT FILING DATE: 2001-06-29

; NUMBER OF SEQ ID NOS: 990

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 573

; LENGTH: 132

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-895-814-573

Alignment Scores:

Pred. No.: 2,05e-19

Length: 132

Score:	286.00	Matches:	60
Percent Similarity:	65.42%	Conservative:	10
Percent Local Similarity:	56.07%	Mismatches:	33
Query Match:	2.97%	Indels:	5
DB:	9	Gaps:	2

US-09-966-880A-9 (1-5514) x US-09-895-814-573 (1-132)

3851	QY	AA	TTTTTCTCTCTTATATTTTTTTTGGAGCAAGGCTCGTGCCCTGTACCCAGGCTGGAGTGCAG	3792
			:::	
27	Db	Asp	Phe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln	44
3791	QY	TG	CGCGCAAGCACGTGCTCAGTGCAGCGCTTGACCTCCTGGGCTCCAAAGATTCTCCTGCCTC	3732
			:::	
45	Db	Trp	HisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu	64
3731	QY	AG	CCC--CCTAGTAGCTGGGACCAAGGCACATGCCACCATGCCGGCTAATTTTGT---	3676
			:::	
65	Db	Ser	LeuProHisSerTrpaspPHisArgTyrAlaProProHisLeuAlaAsnPheCysSer	84
3675	QY	--	TATAGAAACAGGGTTTTGCCATGTGTGCTCAGGCTGGTCTTGAACCTCAGGGCTCAAG	3619
			:::	
85	Db	Phe	SerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln	104
3618	QY	CG	ATCCACCACCTCAGCCTCCAGAGTGTCTGGATTACAGCGCTGACACACACACCTG	3559
			:::	
105	Db	Gln	SerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProProHisPro	124
3558	QY	GC	TGTGTGTTTCACTTTAAAT	3538
125	Db	Ala	CysHisIleLeuLeuAsn	131

RESULT 9

```

US-09-759-143-573
; Sequence 573, Application US/09759143
; Patent No. US2002002248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS OF PROSTATE CANCER
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-573

```

Alignment Scores:	
pred. No.:	2,05e-19
Score:	286.00
Percent Similarity:	65.42%
Best Local Similarity:	56.07%
Query Match:	2.97%
Length:	132
Matches:	60
Conservative:	10
Mismatches:	33
Indels:	5

DB:	10	Gaps:	2
US-09-966-880A-9 (1-5514) x US-09-759-143-573 (1-132)			

Qy	3851	AAATTTCTCTCTTATATTTTGTGAGCAAGAGTCTGGGCTCTGCACCCAGGCTGAGGTGCAG	37392
		: : : : : : : : :	
Db	27	AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln	44
Qy	3791	TGCGCAAGCACCTGCAGTGCAGCGCTTGACCTCCCTGGGCTCCAAAAGATTCTCCTGCCTC	37332
		: : : : : : : : :	
Db	45	TripHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu	64
Qy	3731	AGCCG--CCTAGTAGCTGGGACCAAGCAGCATGCCACCATGCCCGCTAATTTTGT---	3676
		: : : : : : : : :	
Db	65	SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer	84
Qy	3675	---TATAGAAACAGGGTTTGGCATGTGTCTCAGGCTGGTCTTGAATCCAGGCGCTCAAG	3619
		: : : : : : : : : : : : : :	
Db	85	PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln	104
Qy	3618	CGATCCACCACCTCAGCCTCCACAGAGTGTGGATTACAGGCTGAGGCACACACCTG	3559
		: : : : : : : : : : : : : :	
Db	105	GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProProHisPro	124
Qy	3558	GCTGCTCTGTTCCTACTTTAAAT	3538
		: : : : : : : : :	
Db	125	AlaCysHisIleLeuLeuAsn	131

RESULT 10

```

US-09-780-669-573
; Sequence 573, Application US/09780669
; Patent No. US20020051977A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
;
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITL OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
;
; FILE REFERENCE: 210121.427C24
;
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
;
; NUMBER OF SEQ ID NOS: 943
;
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
;
; ORGANISM: Homo sapiens
;
US-09-780-669-573

```

Alignment Scores:	
Pred. No.:	2,05e-19
Score:	286.00
Percent Similarity:	65.42%
Best Local Similarity:	56.07%
Query Match:	2.97%
DB:	10
Length:	132
Matches:	60
Conservative:	10
Mismatches:	33
Indels:	5
Gaps:	2

```
US-09-966-880A-9 (1-5514) x US-09-780-669-573 (1-132)
QY 3851 AATTTTCTCTTATTTTGGGCAAGGCTGCTGCTCCTCACCAGGCTGGAGTCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGGGCAAGCAGCTCAGTCAGCGCTTGACCTCCTGGGCTCCAAAGATTCTCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCC-CCTAGTAGCTGGGACACAGGCACATGCCACCATGCCCGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspPheHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATAGAAACAGGGTTTTGCCATGTTGCTCAGGCTGGCTTGAACCTCCAGGCTCAAG 3619
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCACCTCAGCTCCAGAGTGTGGGATTACAGGCGTGAGCCACACCATCG 3559
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
QY 3558 GCTGCTCTGTTCACTTAAT 3538
Db 125 AlaCysHisIleLeuLeuAsn 131
RESULT 11
US-09-822-827-573
; Sequence 573, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-827-573
Alignment Scores:
Pred. No.: 2,05e-19 Length: 132
Score: 286.00 Matches: 60
Percent Similarity: 65.42% Conservative: 10
Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: 10 Gaps: 2
US-09-966-880A-9 (1-5514) x US-09-822-827-573 (1-132)
QY 3851 AATTTTCTCTTATTTTGGGCAAGGCTGCTGCTCCTCACCAGGCTGGAGTCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGGGCAAGCAGCTCAGTCAGCGCTTGACCTCCTGGGCTCCAAAGATTCTCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCC-CCTAGTAGCTGGGACACAGGCACATGCCACCATGCCCGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspPheHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATAGAAACAGGGTTTTGCCATGTTGCTCAGGCTGGCTTGAACCTCCAGGCTCAAG 3619
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCACCTCAGCTCCAGAGTGTGGGATTACAGGCGTGAGCCACACCATCG 3559
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
QY 3558 GCTGCTCTGTTCACTTAAT 3538
Db 125 AlaCysHisIleLeuLeuAsn 131
RESULT 12
US-10-012-896-884
; Sequence 884, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriack
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 884
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-884
Alignment Scores:
Pred. No.: 2,07e-19 Length: 135
Score: 286.00 Matches: 60
Percent Similarity: 65.42% Conservative: 10
Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: 9 Gaps: 2
US-09-966-880A-9 (1-5514) x US-10-012-896-884 (1-135)
QY 3851 AATTTTCTCTTATTTTGGGCAAGGCTGCTGCTCCTCACCAGGCTGGAGTCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGGGCAAGCAGCTCAGTCAGCGCTTGACCTCCTGGGCTCCAAAGATTCTCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCC-CCTAGTAGCTGGGACACAGGCACATGCCACCATGCCCGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspPheHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATAGAAACAGGGTTTTGCCATGTTGCTCAGGCTGGCTTGAACCTCCAGGCTCAAG 3619
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
```

Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
 Qy 3618 CGATCCACCCACCTCCAGCCTCCAGAGTGTGGAGTACAGCGTGCAGCCACACACCTG 3559
 Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
 Qy 3558 GCTGCTGCTTTCACCTTTAAAT 3538
 Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 13

US-09-895-793-884
 ; Sequence 884, Application US/09895793
 ; Publication No. US20020192763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, JIANGCHUN
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Vinals de Bassols, Carlota
 ; APPLICANT: Foy, Teresa
 ; APPLICANT: Fanger, Gary R.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.534C2
 ; CURRENT APPLICATION NUMBER: US/09/895,793
 ; CURRENT FILING DATE: 2001-06-29
 ; NUMBER OF SEQ ID NOS: 982
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 884
 ; LENGTH: 135
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-895-793-884

Alignment Scores:
 Pred. No.: 2,07e-19 Length: 135
 Score: 286.00 Matches: 60
 Percent Similarity: 65.42% Conservative: 10
 Best Local Similarity: 56.07% Mismatches: 33
 Query Match: 9 Indels: 5
 DB: Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-895-793-884 (1-135)

Qy 3851 AATTTTCTCTTATTTTGTGAGCAAGGTCTGGTCTGTACCCAGGCTGGAGTGCAG 3792
 Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
 Qy 3791 TGGCGCAAGCACTGCTCAGTCAGCCTTGACCTCTGGCTCCAAAGATTCTCTGCTC 3732
 Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
 Qy 3731 AGCCC-CCTAGTAGCTGGACACACGACATGCCACCATGCCCGCGCTAATTTTGT- 3676
 Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProHisLeuAlaAsnPheCysSer 84
 Qy 3675 ---TATAGAAACAGGGTTTGGCCATGTTGCTCAGGCTGGTCTTGAACATCCAGGCTCAAG 3619

Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
 Qy 3618 CGATCCACCCACCTCCAGCCTCCAGAGTGTGGAGTACAGCGTGCAGCCACACACCTG 3559
 Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
 Qy 3558 GCTGCTGCTTTCACCTTTAAAT 3538
 Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 14

US-09-895-814-884
 ; Sequence 884, Application US/09895814
 ; Publication No. US20020193296A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, JIANGCHUN
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel X.
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Hural, John
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Vinals de Bassols, Carlota
 ; APPLICANT: Foy, Teresa
 ; APPLICANT: Fanger, Gary R.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.427C26
 ; CURRENT APPLICATION NUMBER: US/09/895,814
 ; CURRENT FILING DATE: 2001-06-29
 ; NUMBER OF SEQ ID NOS: 990
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 884
 ; LENGTH: 135
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-895-814-884

Alignment Scores:
 Pred. No.: 2,07e-19 Length: 135
 Score: 286.00 Matches: 60
 Percent Similarity: 65.42% Conservative: 10
 Best Local Similarity: 56.07% Mismatches: 33
 Query Match: 9 Indels: 5
 DB: Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-895-814-884 (1-135)

Qy 3851 AATTTTCTCTTATTTTGTGAGCAAGGTCTGGTCTGTACCCAGGCTGGAGTGCAG 3792
 Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
 Qy 3791 TGGCGCAAGCACTGCTCAGTCAGCCTTGACCTCTGGCTCCAAAGATTCTCTGCTC 3732
 Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
 Qy 3731 AGCCC-CCTAGTAGCTGGACACACGACATGCCACCATGCCCGCGCTAATTTTGT- 3676
 Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProHisLeuAlaAsnPheCysSer 84

QY 3675 ---TATAGAAACAGGGTTTCCCATGTTGCTCAGCGTGGTCTTGAACCTCCAGGCTCAAG 3619
 Db 85 PheSerArgAspGlyValSerLeuCysSerGlyTrpSerLysThrProGlyLeuGln 104
 QY 3618 CGATCCACCCACCTCCAGAGTGTGGGATTACAGCGCTGAGCCACACACCTG 3559
 Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProProHisPro 124
 QY 3558 GCTGCTCTGTTTCACTTTAAAT 3538
 Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 15

US-09-759-143-884
 ; Sequence 884, Application US/09759143
 ; Patent No. US2002002248A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuqi
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.427C23
 ; CURRENT APPLICATION NUMBER: US/09/759,143
 ; NUMBER OF SEQ ID NOS: 934
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 884
 ; LENGTH: 135
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-759-143-884

Alignment Scores:
 Pred. No.: 2,07e-19 Length: 135
 Score: 286.00 Matches: 60
 Percent Similarity: 65.42% Conservative: 10
 Best Local Similarity: 56.07% Mismatches: 33
 Query Match: 2.97% Indels: 5
 DB: 10 Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-759-143-884 (1-135)

QY 3851 AATTTTCTCTCTATTTTGGCAAGTCTGCTCTCACCACCCAGGCTGGAGTGCAG 3792
 Db 27 AsnPhe-----PhePheLeuArgGlnGlnSerValAlaGlnAlaGlyValGln 44
 QY 3791 TGGCGCAGCAGCTGCTCAGTCAGCCTGACCTCTGGGCTCCAAAGATTCTCCTGCCTC 3732
 Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
 QY 3731 AGCCC-CCTAGTAGTCGACGACGACATGCCACCATGCCCGGCTAATTTTGT--- 3676
 Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProHisLeuAlaAsnPheCysSer 84
 QY 3675 ---TATAGAAACAGGGTTTCCCATGTTGCTCAGCGTGGTCTTGAACCTCCAGGCTCAAG 3619
 Db 85 PheSerArgAspGlyValSerLeuCysSerGlyTrpSerLysThrProGlyLeuGln 104

QY 3618 CGATCCACCCACCTCAGCCTCCAGAGTGTGGGATTACAGCGCTGAGCCACACACCTG 3559
 Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProProHisPro 124
 QY 3558 GCTGCTCTGTTTCACTTTAAAT 3538
 Db 125 AlaCysHisIleLeuLeuAsn 131

Search completed: June 14, 2003, 19:09:31
 Job time : 191.542 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 11:14:11 ; Search time 13758 Seconds
(without alignments)
11663.968 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacatcgtgtcca.....tcaaactcctgacctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4771	86.5	11204	9	AB040430 Homo sapi
2	4649	84.3	71132	9	AC092184 Homo sapi
3	84	1.5	1837	9	BC006296 Homo sapi
4	84	1.5	2791	9	AB040431 Homo sapi
c 5	75	1.4	84778	9	AL353650 Human DNA
c 6	75	1.4	123253	9	AC004971 Homo sapi
7	75	1.4	168524	2	AL356418 Homo sapi
8	74	1.3	29359	2	AP001083 Homo sapi
9	74	1.3	168079	2	AC024709 Homo sapi
10	74	1.3	172376	2	AL353749 Homo sapi
11	74	1.3	193280	9	AL139113 Human DNA
c 12	74	1.3	215719	9	AC048334 Homo sapi
c 13	73	1.3	173943	2	AC022717 Homo sapi
14	73	1.3	195661	2	AC027595 Homo sapi
15	73	1.3	214696	2	AF235101 Homo sapi
c 16	71	1.3	152129	2	AC027416 Homo sapi
c 17	71	1.3	206854	9	AC009974 Homo sapi
18	69	1.3	3187	9	AK092318 Homo sapi
19	69	1.3	79419	9	HSJ324017 Human DNA
20	69	1.3	81364	2	AC021072_3 Continuation (4 of
21	69	1.3	109138	9	AC104084 Homo sapi
c 22	69	1.3	127862	9	AL445430 Human DNA
23	69	1.3	146574	2	AC023459 Homo sapi
24	69	1.3	154162	2	AC069270 Homo sapi
c 25	69	1.3	157807	9	AC073573 Homo sapi
c 26	69	1.3	164168	9	AL135927 Human DNA
27	69	1.3	164179	9	AC007227 Homo sapi
c 28	69	1.3	193717	2	AC068888 Homo sapi
c 29	68	1.2	4285	9	AC093077 Homo sapi
30	68	1.2	8429	9	AF350881 Homo sapi
31	68	1.2	39212	9	HSN21F1 Z94162 Human DNA s
c 32	68	1.2	44890	9	AC005778 Homo sapi
c 33	68	1.2	51012	9	AC025536 Homo sapi
34	68	1.2	57779	2	AC115982 Homo sapi
35	68	1.2	90543	9	AC090287 Homo sapi
36	68	1.2	111035	9	AC022405 Homo sapi
c 37	68	1.2	131928	9	AC091738 Homo sapi
c 38	68	1.2	135693	9	HSJ1167H4 AL121914 Human DNA
c 39	68	1.2	141655	9	HS1093G12 AL121751 Human DNA
c 40	68	1.2	146568	2	AC022651 Homo sapi
c 41	68	1.2	148103	9	AL445232 Human DNA
c 42	68	1.2	149726	9	AC016394 Homo sapi
c 43	68	1.2	150192	9	AC008686 Homo sapi
c 44	68	1.2	156949	9	AL161652 Human DNA
45	68	1.2	157860	9	AL391241 Human DNA

ALIGNMENTS

RESULT 1
AB040430 Homo sapiens AID gene for activation-induced cytidine deaminase, complete cds.
LOCUS AB040430
DEFINITION Homo sapiens AID gene for activation-induced cytidine deaminase, complete cds.
ACCESSION AB040430.1 GI:9988407
VERSION AB040430.1
KEYWORDS AID; activation-induced cytidine deaminase.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Muto,T., Muramatsu,M., Tanikawa,K. and Honjo,T.
TITLE Isolation, tissue distribution, and chromosomal localization of the

JOURNAL MEDLINE REFERENCE AUTHORS	human activation-induced cytidine deaminase (AID) gene Genomics 68 (1), 85-88 (2000) 2 (sites) Revu.P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O., Catalan,N., Forveille,M., Dufourcq-Lagelouse,R., Gennery,A., Tezcan,I., Ersoy,F., Kayserili,H., Ugazio,A.G., Brousse,N., Muramatsu,M., Taniguchi,K., Kinoshita,K., Honjo,T., Fischer,A. and Durandy,A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2) Cell 102 (5), 565-575 (2000) 20460541	Db	421	AATGCACCTGTACAGCTGAGACAGAGAACCATCAATTAATTGAAGTGTGAGATTTTCTGGCCT	480
JOURNAL MEDLINE REFERENCE	3 (bases 1 to 11204) Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T. Direct Submission Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail:honjoemfour.med.kyoto-u.ac.jp, Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388) Location/Qualifiers 1. .11204 /organism="Homo sapiens" /db_xref="taxon:9606" Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009) /gene="AID" Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009) /gene="AID" /codon_start=1 /product="activation-induced cytidine deaminase" /protein_id="BAB1270.1" /db_xref="GI:9988408" /translation="MDSLILMMRRKFLYQFNVRWAKGRRETYLCYVVRKDSATSFSLDFGLRNKGVHVELFLRYISDMDLDFGRVYRWFTWSPCYDCARHVDLFRGNPNLSLRIFFARLYFCDDRKAEPGLRLRLHRAVGIOIAIMFKDYFCWNTFVENHRETEKAWEGHENSVRSLRLLPLLYEVDLDRFAFTLGL" BASE COUNT 3305 a 2273 c 2373 g 3253 t ORIGIN	QY	1071	GAGACTTGCAGGAGGCAAGACACACTCTGGACACCACTATATGACAGGTAAGAGGCGAG	1130
JOURNAL MEDLINE REFERENCE	1071	Db	481	GAGACTTGCAGGAGGCAAGACACACTCTGGACACCACTATATGACAGGTAAGAGGCGAG	540
JOURNAL MEDLINE REFERENCE	1131	QY	1131	TCTTCGCGTGGGTGATTTGGCACTGGCCTTCTCTCAGAGCAAAATCTCAGTAATCAGACTGG	1190
JOURNAL MEDLINE REFERENCE	541	Db	541	TCTTCGCGTGGGTGATTTGGCACTGGCCTTCTCTCAGAGCAAAATCTCAGTAATCAGACTGG	600
JOURNAL MEDLINE REFERENCE	1191	QY	1191	TAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAGACTGAGCTTGTATCAATATGCA	1250
JOURNAL MEDLINE REFERENCE	601	Db	601	TAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAGACTGAGCTTGTATCAATATGCA	660
JOURNAL MEDLINE REFERENCE	1251	QY	1251	TATATATTTTGTGATCTGCTCCTCTTCTTATTCAGATCTTATACGGTGTGAGCCCAA	1310
JOURNAL MEDLINE REFERENCE	661	Db	661	TATATATTTTGTGATCTGCTCCTCTTCTTATTCAGATCTTATACGGTGTGAGCCCAA	720
JOURNAL MEDLINE REFERENCE	1311	QY	1311	TTCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGCAAAAGAGTAGTG	1370
JOURNAL MEDLINE REFERENCE	721	Db	721	TTCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGCAAAAGAGTAGTG	780
JOURNAL MEDLINE REFERENCE	1371	QY	1371	CGTACAACTACTGATTCGCTCTGAGATTTGTACCATGGTTGAACTAATTTATGTAAT	1430
JOURNAL MEDLINE REFERENCE	781	Db	781	CGTACAACTACTGATTCGCTCTGAGATTTGTACCATGGTTGAACTAATTTATGTAAT	840
JOURNAL MEDLINE REFERENCE	1431	QY	1431	AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTA	1490
JOURNAL MEDLINE REFERENCE	841	Db	841	AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTA	900
JOURNAL MEDLINE REFERENCE	1491	QY	1491	CTAAACCGGTAGTGTCTAATTTTTCGTAATTAATTTTCTAATATTTCAACAGTAAACAACAT	1550
JOURNAL MEDLINE REFERENCE	901	Db	901	CTAAACCGGTAGTGTCTAATTTTTCGTAATTAATTTTCTAATATTTCAACAGTAAACAACAT	960
JOURNAL MEDLINE REFERENCE	1551	QY	1551	TGAAGACACACTTTCCTAGGAGGCGTCTACTGAAATAATTTAGCTATAGTAAGAAATTT	1610
JOURNAL MEDLINE REFERENCE	961	Db	961	TGAAGACACACTTTCCTAGGAGGCGTCTACTGAAATAATTTAGCTATAGTAAGAAATTT	1020
JOURNAL MEDLINE REFERENCE	1611	QY	1611	GTAAATTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTTGAAAGTCACATATGATGTGTC	1670
JOURNAL MEDLINE REFERENCE	1021	Db	1021	GTAAATTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTTGAAAGTCACATATGATGTGTC	1080
JOURNAL MEDLINE REFERENCE	1671	QY	1671	CATTATAGGAGACAAATTCATTCAGCAAGTATTAATTTTAAAGGCCCAATTTGTTAG	1730
JOURNAL MEDLINE REFERENCE	1081	Db	1081	CATTATAGGAGACAAATTCATTCAGCAAGTATTAATTTTAAAGGCCCAATTTGTTAG	1140
JOURNAL MEDLINE REFERENCE	1731	QY	1731	GCAGTTAATGGCACATTTTACTTAATTAATCTTTCCATTTGTTTCAGAGCTAGCTTAACCT	1790
JOURNAL MEDLINE REFERENCE	1141	Db	1141	GCAGTTAATGGCACATTTTACTTAATTAATCTTTCCATTTGTTTCAGAGCTAGCTTAACCT	1200
JOURNAL MEDLINE REFERENCE	1791	QY	1791	TACCTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAAATCTTTATGTGCAGTTTTTGTAT	1850
JOURNAL MEDLINE REFERENCE	1201	Db	1201	TACCTCTTAGGTGTGAATTTGGTTAAGGTCCTCATAAATCTTTATGTGCAGTTTTTGTAT	1260
JOURNAL MEDLINE REFERENCE	1851	QY	1851	AGGTTATTTGTCATAGAACTTATCTATCTTCTACATTTTATGATTTACTATGGATGTATGAGA	1910
JOURNAL MEDLINE REFERENCE	1261	Db	1261	AGGTTATTTGTCATAGAACTTATCTATCTTCTACATTTTATGATTTACTATGGATGTATGAGA	1320
JOURNAL MEDLINE REFERENCE	1911	QY	1911	ATAACACCTTAATCTCTTATACCTCAATTTTAACTCCCTTTTATAAAGAACTTACATTAAC	1970
JOURNAL MEDLINE REFERENCE	1321	Db	1321	ATAACACCTTAATCTCTTATACCTCAATTTTAACTCCCTTTTATAAAGAACTTACATTAAC	1380
JOURNAL MEDLINE REFERENCE	1971	QY	1971	AGAATAAAGATTTTTTAAATAATATATTTTTTGTAGAGACAGGGCTTTAGCCCGCCGAG	2030
JOURNAL MEDLINE REFERENCE	1381	Db	1381	AGAATAAAGATTTTTTAAATAATATATTTTTTGTAGAGACAGGGCTTTAGCCCGCCGAG	1440
JOURNAL MEDLINE REFERENCE	2031	QY	2031	GCTGCTCTTAAGTCTTGCCCAAGGATCTCTGCTGCGGCTCCCTAAAGCTGCTGGA	2090
JOURNAL MEDLINE REFERENCE	1441	Db	1441	GCTGCTCTTAAGTCTTGCCCAAGGATCTCTGCTGCGGCTCCCTAAAGCTGCTGGA	1500
JOURNAL MEDLINE REFERENCE	2091	QY	2091	TTATAGACATGAGCCATCACATCCCAATATACAGATAAAGATTTTTTAAATGAGGATTTAA	2150
JOURNAL MEDLINE REFERENCE	1501	Db	1501	TTATAGACATGAGCCATCACATCCCAATATACAGATAAAGATTTTTTAAATGAGGATTTAA	1560

|||||
38335 TCTCCACAAAAATGGGAAAAAAGAAACCAATCAGTGGTGTCTCTGTGGGAGGGGAAG 38394
QY 4441 GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGTGGGATTCAGGTTCCTGTATCCT 4500
Db 38395 GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGTGGGATTCAGGTTCCTGTATCCT 38454
QY 4501 GACTGTGTAGACAGTTTGGGGTGTTCATCATCCAAAAATATTCGTAGAAATTCATCATCTTA 4560
Db 38455 GACTGTGTAGACAGTTTGGGGTGTTCATCATCCAAAAATATTCGTAGAAATTCATCATCTTA 38514
QY 4561 AATGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAAATAATGTGTAAG 4620
Db 38515 AATGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAAATAATGTGTAAG 38574
QY 4621 AAAAGTTTCAATCTCTTCGCCAGCAACAGTTATTCAAATTCCTGGAGCCCTTACTTCGCA 4680
Db 38575 AAAACTTCAATCTCTTCGCCAGCAACAGTTATTCAAATTCCTGGAGCCCTTACTTCGCA 38634
QY 4681 AATTCTCTGCACCTTCTGCCCGTACCATTAGGTGACAGCACTAGTCCACAAAATTCGATA 4740
Db 38635 AATTCTCTGCACCTTCTGCCCGTACCATTAGGTGACAGCACTAGTCCACAAAATTCGATA 38694
QY 4741 AATGCAATTCCTGGAAGAGACTAGGACAAAATCCAGGCATCCTGTGCTTTTCATATCAA 4800
Db 38695 AATGCAATTCCTGGAAGAGACTAGGACAAAATCCAGGCATCCTGTGCTTTTCATATCAA 38754
QY 4801 CCAGCGTGTACAGCTTGTGTCTGCTGCAATGCAAGTGGGACACTTCATTTCTTTAA 4860
Db 38755 CCATGCTGTACAGCTTGTGTCTGCTGCAATGCAAGTGGGACACTTCATTTCTTTAA 38814
QY 4861 GGAACCTGGGTTACCAGAGTATTTCCACAAATGCTATTCAAAATAGTGCCTATGATATG 4920
Db 38815 GGAACCTGGGTTACCAGAGTATTTCCACAAATGCTATTCAAAATAGTGCCTATGATATG 38874
QY 4921 CAAGACACTGTCTAGAGCCAGAAACAAAGAGGAGGAAATCAGTCATTTATGTGGGA 4980
Db 38875 CAAGACACTGTCTAGAGCCAGAAACAAAGAGGAGGAAATCAGTCATTTATGTGGGA 38934
QY 4981 ACAACATAGCAAGATATTTAGATCATTTTCACTAGTTTAAAGAGCAGCAGTACAAAAT 5040
Db 38935 ACAACATAGCAAGATATTTAGATCATTTTCACTAGTTTAAAGAGCAGCAGTACAAAAT 38994
QY 5041 CACACATGCAATCAGTATATCCAAATCATGTAATATGCTGTGCTGTAGAAAGACTAGAGG 5100
Db 38995 CACACATGCAATCAGTATATCCAAATCATGTAATATGCTGTGCTGTAGAAAGACTAGAGG 39054
QY 5101 AATAACAACAAGAAATCTTAACAGTCAATGCTATTAGACACTAAGTCAATATTATTATT 5160
Db 39055 AATAACAACAAGAAATCTTAACAGTCAATGCTATTAGACACTAAGTCAATATTATTATT 39114
QY 5161 AGACACTATGATATTGAGATTTAAANAATCTTTAATATTTTAAATTTAGAGCTCTTCT 5220
Db 39115 AGACACTATGATATTGAGATTTAAANAATCTTTAATATTTTAAATTTAGAGCTCTTCT 39174
QY 5221 ATTTTTCATAGTATTCAGTTTGAATGATGATCAAGTATTAATCTCTCTTTTCTTTT 5280
Db 39175 ATTTTTCATAGTATTCAGTTTGAATGATGATCAAGTATTAATCTCTCTTTTCTTTT 39234
QY 5281 TTTTCTTTTCTTTTGGAGTGGAGTTTGGTCTTTGTTGCCCTGCTGAGTGGAGTGGCATG 5340
Db 39235 TTTTCTTTTCTTTTGGAGTGGAGTTTGGTCTTTGTTGCCCTGCTGAGTGGAGTGGCATG 39294
QY 5341 AYCATTAGTCTCACTGCAACCTCCACCTCTCTGGTGTCAAGCAAGCTGTGCTCAGCCPTCC 5400
Db 39295 ACCATTAGTCTCACTGCAACCTCCACCTCTCTGGTGTCAAGCAAGCTGTGCTCAGCCPTCC 39354
QY 5401 CGGTAGATGGGATTTACAGGGGCCCCACCACCACTCGGCTAAGTCTTTGTTATTTTACTA 5460
Db 39355 CGGTAGATGGGATTTACAGGGGCCCCACCACCACTCGGCTAAGTCTTTGTTATTTTACTA 39414
QY 5461 GAGATGGGGTTTACCATTGTTGGCAGGCTGGTCTCAAACTCCTGACCTCAGAG 5514
|||||

Db 39415 GAGATGGGGTTTACCATGTTGCCAGGCTGTCTCAAACTCCTGACCTCAGAG 39468
RESULT 3
LOCUS BC006296 1837 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, activation-induced cytidine deaminase, clone
MGC:12911 IMAGE:4054915, mRNA, complete cds.
ACCESSION BC006296
VERSION BC006296.1 GI:13623400
KEYWORDS MGC
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1837)
AUTHORS Strausberg, R.
TITLE Direct Submision
JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Louis Staudt
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland:
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nigri.nih.gov
Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 17 Row: a Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9988409.

FEATURES
source

1..1837
/organism="Homo sapiens"
/db_xref="LocusID:57379"
/db_xref="taxon:9606"
/clone="MGC:12911 IMAGE:4054915"
/tissue_type="Primary B-Cells from Tonsils"
/clone_lib="NIH_MGC_48"
/lab_host="DH10B-R"
/notes="vector: pOTB7"
86..682
/codon_start=1
/product="activation-induced cytidine deaminase"
/protein_id="AAH06296.1"
/db_xref="GI:13623401"
/translation="MDSLLMNRKFLYQFKNRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELFLRYISDMDLDPGRCYRVTFWTSNSPCYDCARHVAFLRGNP
NLSURIFATRLYFCEDRKARPEGLRLRRAGVQIAIMTKDYFCWNTFVENHRTFK
AWELGHSNRYLSRLRILLPLLEVDLDRQAFRTGLI"
BASE COUNT 530 a 387 c 421 g 499 t

CDS

Query Match
Best Local Similarity 100.0%; Score 84; DB 9; Length 1837;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1035 GAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGA 1094

```

|||||
10 GAACCATCATTAATGAAGTGGAGATTTTCGGCCGTGAGACTTGCGAGGAGGCAAGAAGA 69
|||||
OY 1095 CACTCTGGACACCACTATGGACAG 1118
|||||
Db 70 CACTCTGGACACCACTATGGACAG 93
|||||

RESULT 4
AB040431
LOCUS 2791 bp mRNA linear PRI 03-OCT-2000
DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,
complete CDS.
ACCESSION AB040431
VERSION AB040431.1 GI:9988409
KEYWORDS AID; activation-induced cytidine deaminase; Human AID.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE Isolation, tissue distribution, and chromosomal localization of the
human activation-induced cytidine deaminase (AID) gene
JOURNAL Genomics 68 (1), 85-88 (2000)
MEDLINE 20408890
REFERENCE 2 (sites)
Revy,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O.,
Catalan,N., Forveille,M., Dufourcq-Lagelouse,R., Gennery,A.,
Tescan,I., Ersoy,F., Kayserili,H., Ugazio,A.G., Brenner,N.,
Muramatsu,M., Notarangelo,L.D., Kinoshita,K., Honjo,T., Fischer,A.
and Durandy,A.
TITLE Activation-induced cytidine deaminase (AID) deficiency causes the
autosomal recessive form of the Hyper-IgM syndrome (HIGM2)
JOURNAL Cell 102 (5), 565-575 (2000)
MEDLINE 20460541
REFERENCE 3 (bases 1 to 2791)
Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
Direct Submission
TITLE Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department
of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku,
Kyoto, Kyoto 606-8501, Japan (E-mail:honjoemfour.med.kyoto-u.ac.jp,
Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)
JOURNAL Location/Qualifiers
FEATURES
Source 1. 2791
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .2791
/gene="AID"
77. .673
/gene="AID"
/codon_start=1
/product="activation-induced cytidine deaminase"
/protein_id="BAB12721.1"
/db_xref="GI:9988410"
/translaton="MDSLMMNRKFLYQKVRWAKGRRETYLCVVKRRDSATSFSL
DFGYLRKNGCHVELFLRYISDMWLDPCRCYRVTFWFSWSPCYDCARHVADFLGNP
NLSRIFTRILYFQEDRKAPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTKF
AWEGHENSVRSLRLLPLLYEVDLDRDAFRTLGL"
BASE COUNT 842 a 548 c 625 g 776 t
ORIGIN
Query Match 1.5%; Score 84; DB 9; Length 2791;
Best Local Similarity 100.08; Pred. No. 6.9e-34;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1035 GAACCATCATTAATGAAGTGGAGATTTTCGGCCGTGAGACTTGCGAGGAGGCAAGAAGA 1094
|||||
Db 1 GAACCATCATTAATGAAGTGGAGATTTTCGGCCGTGAGACTTGCGAGGAGGCAAGAAGA 60
|||||
OY 1095 CACTCTGGACACCACTATGGACAG 1118
|||||
Db 61 CACTCTGGACACCACTATGGACAG 84
|||||

```

RESULT 5
AL353650/c
LOCUS 84778 bp DNA linear PRI 19-SEP-2001
DEFINITION Human DNA sequence from clone RP11-344N19 on chromosome 10.
Contains STSS, GSSs and a CpG island, complete sequence.
ACCESSION AL353650
VERSION AL353650.5 GI:14141281
KEYWORDS HTG; CpG island.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 84778)
Griffiths,C.
Direct Submission
Submitted (01-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 17, 2001 this sequence version replaced gi:13567986.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TRMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-344N19 is from the library RPC1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-344N19 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-344N19 is at 1 in this sequence.
The true left end of clone RP11-735G18 is at 84679 in this
sequence. The true right end of clone RP11-426A21 is at 3263 in
this sequence.

FEATURES
Source
1. 84778
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-344N19"
/clone_lib="RPC1-11.2"
14. .616
/note="match: GSS: Em:AQ542535"
102. .282
/note="LIP1A7 repeat: matches 5960. .6141 of consensus"
1753. .2298
/note="LIP1A14 repeat: matches 5601. .6149 of consensus"
4213. 4511
/note="AluX repeat: matches 1. .300 of consensus"
4643. .4720
/note="MIR repeat: matches 168. .245 of consensus"
4733. .5068
/note="MER2 repeat: matches 8. .345 of consensus"

misc_feature
14. .616
/note="match: GSS: Em:AQ542535"
102. .282
/note="LIP1A7 repeat: matches 5960. .6141 of consensus"
1753. .2298
/note="LIP1A14 repeat: matches 5601. .6149 of consensus"
4213. 4511
/note="AluX repeat: matches 1. .300 of consensus"
4643. .4720
/note="MIR repeat: matches 168. .245 of consensus"
4733. .5068
/note="MER2 repeat: matches 8. .345 of consensus"


```
repeat_region 5314. .5450
/note="L2 repeat: matches 2552. .2710 of consensus"
repeat_region 7002. .7298
/note="AluX repeat: matches 1. .297 of consensus"
misc_feature complement(7572. .7912)
/note="match: GSS: Em:AQ079921"
misc_feature 7685. .8099
/note="match: GSS: Em:AQ569094"
repeat_region 8242. .8529
/note="AluJo repeat: matches 1. .286 of consensus"
repeat_region 8905. .9048
/note="MIR repeat: matches 16. .161 of consensus"
repeat_region 9051. .9181
/note="FLAM_C repeat: matches 1. .129 of consensus"
repeat_region 9183. .9233
/note="MIR repeat: matches 154. .206 of consensus"
repeat_region 9381. .9442
/note="L2 repeat: matches 2593. .2654 of consensus"
repeat_region 9439. .9568
/note="MIR repeat: matches 39. .171 of consensus"
repeat_region 9760. .9934
/note="match: GSS: Em:AQ038279"
misc_feature 10185. .10572
complement(10271. .10734)
/note="match: GSS: Em:AQ458578"
misc_feature 10749. .11320
/note="match: GSS: Em:AQ413467"
repeat_region 10843. .10888
/note="23 copies 2 mer aa 76% conserved"
misc_feature complement(11059. .11336)
/note="match: STS: Em:G54666"
misc_feature complement(11100. .11336)
/note="match: STS: Em:G59447"
repeat_region 11331. .11636
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 11703. .11754
/note="MER5B repeat: matches 3. .54 of consensus"
repeat_region 12133. .12491
/note="THEIC repeat: matches 1. .371 of consensus"
repeat_region 12598. .12901
/note="AluX repeat: matches 1. .303 of consensus"
repeat_region 13917. .14143
/note="L1 repeat: matches 3687. .3934 of consensus"
repeat_region 14489. .14780
/note="AluX repeat: matches 1. .295 of consensus"
repeat_region 15575. .15719
/note="FRAM repeat: matches 1. .146 of consensus"
repeat_region 16603. .18432
/note="L1MEC repeat: matches 4. .1865 of consensus"
repeat_region 18433. .18727
/note="AluSg repeat: matches 1. .295 of consensus"
repeat_region 18728. .18931
/note="L1MEC repeat: matches 1865. .2068 of consensus"
repeat_region 18981. .19279
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 19399. .19510
/note="L1MEC repeat: matches 2159. .2268 of consensus"
repeat_region 19511. .20314
/note="L1MA6 repeat: matches 5511. .6300 of consensus"
repeat_region 20315. .20795
/note="L1MEC repeat: matches 2268. .2414 of consensus"
repeat_region 20756. .21203
/note="L1M4 repeat: matches 2375. .2833 of consensus"
repeat_region 21278. .21396
/note="L1PR2 repeat: matches 6036. .6155 of consensus"
repeat_region 21531. .21805
/note="AluX repeat: matches 1. .275 of consensus"
repeat_region 21807. .22256
/note="L1M4 repeat: matches 2462. .2946 of consensus"
repeat_region 22400. .22500
/note="L1M4 repeat: matches 7863. .7964 of consensus"
repeat_region 22631. .22863

Query Match 1.4%; Score 75; DB 9; Length 84778;
Best Local Similarity 100.0%; Pred. No. 4.5e-29;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GCCTAATCTTGTATTTTAGATAGATGGGTTTCCACCATGTTGCCAGGCTGCTCA 5497
|||||
Db 23086 GCCTAATCTTGTATTTTAGATAGATGGGTTTCCACCATGTTGCCAGGCTGCTCA 23027
|||||

QY 5498 AACTCCTGACCTCAG 5512
|||||
Db 23026 AACTCCTGACCTCAG 23012

RESULT 6
```

```

AC004971/c
LOCUS       AC004971      123253 bp      DNA      linear      PRI 30-SEP-2000
DEFINITION  Homo sapiens PAC clone RP5-1125K23 from 7, complete sequence.
ACCESSION   AC004971
VERSION     AC004971.3  GI:10047915
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens.
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1  (bases 1 to 123253)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
  PUBMED    9847074
REFERENCE   2  (bases 1 to 123253)
  AUTHORS   Stromatt,C., Ali,J. and Miller,K.
  TITLE     The sequence of Homo sapiens: PAC clone RP5-1125K23
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 123253)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
            Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 123253)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
            Submitted (10-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 123253)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
            Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Sep 10, 2000 this sequence version replaced gi:7630775.
COMMENT     ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: H_DJ1125K23
            -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by
Pieter de Jong and coworkers at the Roswell Park Cancer Institute

```

(http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-607J2; the clone sequenced to the right is RP5-826E18, 200 base pair overlap. Actual end of this clone is at base position 123057 of RP5-1125K23.

There are polymorphic base pair differences in the overlap between the clone RP5-1125K23 and RP4-607J2.

FEATURES

FEATURES	Location/Qualifiers	source
	1..123253	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7"	
	/clone="RP5-1125K23"	
	/clone_lib="RPCI-5"	
	65..226	
misc_feature	/note="similar to EST AAL58750 (NID:g1733597) zo63e09.r1"	
misc_feature	65..225	
	/note="similar to EST AA079742 (NID:g1618634) zm20g10.r1"	
misc_feature	65..225	
	/note="similar to EST AA486040 (NID:g2216256) ab40c06.r1"	
misc_feature	65..225	
	/note="similar to EST AW579567 (NID:g7254616)"	
misc_feature	65..225	
	/note="similar to EST BE146164 (NID:g8608888)"	
misc_feature	65..210	
	/note="similar to EST AA679866 (NID:g2656333) ac83h04.s1"	
misc_feature	68..225	
	/note="similar to EST AW804699 (NID:g7856569)"	
misc_feature	68..225	
	/note="similar to EST AW900299 (NID:g8064504)"	
misc_feature	77..225	
	/note="similar to EST AW320240 (NID:g6749784) uo16c08.y1"	
repeat_region	325..616	
	/rpt_family="Alu"	
misc_feature	837..965	
	/note="similar to EST AA079742 (NID:g1618634) zm20g10.r1"	
misc_feature	837..965	
	/note="similar to EST AW320240 (NID:g6749784) uo16c08.y1"	
misc_feature	837..965	
	/note="similar to EST AW579567 (NID:g7254616)"	
misc_feature	837..965	
	/note="similar to EST AW900299 (NID:g8064504)"	
misc_feature	837..965	
	/note="similar to EST BE146164 (NID:g8608888)"	
misc_feature	837..959	
	/note="similar to EST AW804699 (NID:g7856569)"	
misc_feature	837..954	
	/note="similar to EST AA486040 (NID:g2216256) ab40c06.r1"	
misc_feature	865..965	
	/note="similar to EST AA914792 (NID:g3054184) vz09c01.r1"	
repeat_region	1113..1425	
	/rpt_family="Alu"	
misc_feature	1470..1527	
	/note="similar to EST AA914792 (NID:g3054184) vz09c01.r1"	
misc_feature	1470..1527	
	/note="similar to EST AW320240 (NID:g6749784) uo16c08.y1"	
misc_feature	1470..1527	
	/note="similar to EST AW579567 (NID:g7254616)"	
misc_feature	1470..1527	
	/note="similar to EST BE146164 (NID:g8608888)"	
misc_feature	1470..1521	
	/note="similar to EST AA079742 (NID:g1618634) zm20g10.r1"	
misc_feature	1470..1509	
	/note="similar to EST AW900299 (NID:g8064504)"	
misc_feature	1831..1952	

misc_feature /note="similar to EST AA233940 (NID:g1855135) zr27a04.r1" 1831..1952

misc_feature /note="similar to EST AW674791 (NID:g7540101) ba59f07.y1" 1833..1953

misc_feature /note="similar to EST AA914792 (NID:g3054184) vz09c01.r1" 1833..1952

misc_feature /note="similar to EST AW320240 (NID:g6749784) uol6c08.y1" 1833..1934

misc_feature /note="similar to EST BE146164 (NID:g8608888)" 1833..1925

repeat_region /note="similar to EST AW579567 (NID:g7254616)" 2205..2327

repeat_region /rpt_family="LI" 2338..2644

repeat_region /rpt_family="Alu" 2656..2822

repeat_region /rpt_family="Alu" 2879..3182

repeat_region /rpt_family="Alu" 3317..3425

misc_feature /rpt_family="Alu" 3695..3887

misc_feature /note="similar to EST AW674791 (NID:g7540101) ba59f07.y1" 3695..3866

misc_feature /note="similar to EST AA233940 (NID:g1855135) zr27a04.r1" 3695..3784

misc_feature /note="similar to EST AW320240 (NID:g6749784) uol6c08.y1" 3739..3887

misc_feature /note="similar to EST AI199495 (NID:g3752101) qi56f08.x1" 3862..3887

repeat_region /note="similar to EST AW732308 (NID:g7632637) bb02e08.y1" 4752..5441

repeat_region /rpt_family="ERV1" 5526..5656

repeat_region /rpt_family="Alu" 5769..6067

repeat_region /rpt_family="Alu" 6154..6190

repeat_region /rpt_family="Alu" 6220..6500

repeat_region /rpt_family="Alu" 7622..7687

repeat_region /rpt_family="MIR" 7779..7961

repeat_region /rpt_family="Alu" 7965..8265

misc_feature /rpt_family="Alu" 9058..10603

misc_feature /note="CpG island (%GC=66.6, o/e=1.00, #CpGs=168)" 9177..9758

misc_feature /note="similar to EST AW732308 (NID:g7632637) bb02e08.y1" 9177..9365

misc_feature /note="similar to EST AW674791 (NID:g7540101) ba59f07.y1" 9177..9363

misc_feature /note="similar to EST AI199495 (NID:g3752101) qi56f08.x1" 9177..9363

Query Match 1.4%; Score 75; DB 9; Length 123253;
Best Local Similarity 100.0%; Pred. No. 4.5e-29;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GCCTAATGTTGTTATTTTAGTAGAGATGGGTTTACCAGTTGGCCAGGCTGGCTCA 5497
|||||

Db 104338 GGCTAATGTTGTTATTTTAGTAGAGATGGGTTTACCAGTTGGCCAGGCTGGCTCA 104279
|||||

QY 5498 AACTCCTGACCTCAG 5512
|||||

Db 104278 AACTCCTGACCTCAG 104264

RESULT 7
AL356418 168524 bp DNA linear HTG 11-JUL-2001
LOCUS AL356418
DEFINITION Homo sapiens chromosome 10 clone RP11-215117, *** SEQUENCING IN

ACCESSION
AL356418
VERSION
GI:13620377
KEYWORDS
HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168524)
Plumb,B
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:9797524.

COMMENT

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA215117
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: piasmld; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 161402 bases at least Q40
Consensus quality: 164147 bases at least Q30
Consensus quality: 165883 bases at least Q20
Insert size: 167324; sum-of-contigs
Insert size: 158470; 5.5% error; agarose-fp
Quality coverage: 4.52x in Q20 bases; sum-of-contigs Quality
coverage: 4.85x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 19496: contig of 19496 bp in length
19497 19596: gap of 100 bp
19597 38508: contig of 18912 bp in length
38509 38608: gap of 100 bp
38609 55774: contig of 17166 bp in length
55775 55974: gap of 100 bp
55975 81803: contig of 25929 bp in length
81804 81903: gap of 100 bp
81904 89760: contig of 7857 bp in length
89761 89860: gap of 100 bp
89861 96400: contig of 6540 bp in length
96401 96500: gap of 100 bp
96501 126138: contig of 29638 bp in length
126139 126238: gap of 100 bp
126239 138525: contig of 12287 bp in length
138526 138625: gap of 100 bp
138626 151528: contig of 12903 bp in length
151529 151628: gap of 100 bp
151629 156089: contig of 4461 bp in length
156090 156189: gap of 100 bp
156190 158536: contig of 2347 bp in length
158537 158636: gap of 100 bp
158637 166329: contig of 7693 bp in length
166330 166429: gap of 100 bp
166430 168524: contig of 2095 bp in length.

FEATURES
Location/Qualifiers
1..168524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-215117"


```

BASE COUNT      7689 a   6357 c   6443 g   8470 t   400 others
ORIGIN
Query Match      1.3%; Score 74; DB 2; Length 29359;
Best Local Similarity 100.0%; Pred. No. 1.6e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GGCATATGTTTGTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGGTCTCA 5497
      |||||||
Db 22056 GGCATATGTTTGTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGGTCTCA 22115
      |||||||
QY 5498 AACTCTCTGACCTCA 5511
      |||||||
Db 22116 AACTCTCTGACCTCA 22129
      |||||||

RESULT 9
AC024709
LOCUS
DEFINITION      Homo sapiens chromosome 3 clone RP11-147G7, WORKING DRAFT SEQUENCE,
19 unordered pieces.
ACCESSION      AC024709
VERSION        AC024709.5 GI:9838092
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        Waterston,R.H.
TITLE          The sequence of Homo sapiens clone
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 168079)
AUTHORS        Waterston,R.H.
TITLE          Direct Submission
JOURNAL        Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT        On Aug 17, 2000 this sequence version replaced gi:7630865.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0147G07
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 96% of reads
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158772 bases at least Q40
Consensus quality: 160752 bases at least Q30
Consensus quality: 162013 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 167609; sum-of-contigs
Quality coverage: 4.90 in Q20 bases; agarose-fp
Quality coverage: 5.01 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1412: contig of 1412 bp in length
* 1513 1512: gap of unknown length
* 1513 3446: contig of 1934 bp in length
* 3447 3546: gap of unknown length
* 3547 5274: contig of 1728 bp in length

```

FEATURES

source

```

1..168079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="3"
/clone="RP11-147G7"
1..1412
/note="assembly_name:Contig6"
1513..3446
/note="assembly_name:Contig7"
3547..5274
/note="assembly_name:Contig8"
5375..6854
/note="assembly_name:Contig9"
6955..8464
/note="assembly_name:Contig10"
8565..10233
/note="assembly_name:Contig11"
10334..11461
/note="assembly_name:Contig12"
clone_end:SP6
vector_side:right
11562..15323
/note="assembly_name:Contig13"
15424..18905
/note="assembly_name:Contig14"
19006..23670
/note="assembly_name:Contig15"
23771..30058
/note="assembly_name:Contig16"
30159..34404
/note="assembly_name:Contig17"
34505..43175
/note="assembly_name:Contig18"
43276..53770
/note="assembly_name:Contig19"
53871..64674
/note="assembly_name:Contig20"
64775..82701
/note="assembly_name:Contig21"
82802..107645

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```
/note="assembly_name:Contig22"
107746. .135351
/note="assembly_name:Contig23"
135452. .168079
/note="assembly_name:Contig24"
BASE COUNT 49957 a 32303 c 31709 g 52284 t 1826 others
ORIGIN

Query Match      1.3%; Score 74; DB 2; Length 168079;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GCGTAATGTTGTTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCA 5497
      |||
Db 122890 GCGTAATGTTGTTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCA 122949
      |||

QY 5498 AACTCCTGACCTCA 5511
      |||
Db 122950 AACTCCTGACCTCA 122963
      |||

RESULT 10
AL353749
LOCUS      172376 bp      DNA      linear      HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-138018, *** SEQUENCING IN
            PROGRESS ***, 26 unordered pieces.
ACCESSION  AL353749
VERSION     AL353749.3 GI:9863661
KEYWORDS   HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Plumb,B.
            1 (bases 1 to 172376)
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9213188.
----- Genomes Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba138018
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 159786 bases at least Q40
Consensus quality: 163877 bases at least Q30
Consensus quality: 166550 bases at least Q20
Insert size: 169876; sum-of-contigs
Insert size: 191077; agarose-fp
Quality coverage: 3.70x in Q20 bases; sum-of-contigs Quality
coverage: 3.43x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 17327: contig of 17327 bp in length
* 17328 17427: gap of 100 bp
* 17428 20692: contig of 3265 bp in length
* 20693 20792: gap of 100 bp
* 20793 23652: contig of 2860 bp in length
* 23653 23752: gap of 100 bp
*
* 23753 25971: contig of 2219 bp in length
* 25972 26071: gap of 100 bp
* 26072 46738: contig of 20667 bp in length
* 46739 46838: gap of 100 bp
* 46839 56508: contig of 9670 bp in length
* 56509 56608: gap of 100 bp
* 56609 59070: contig of 2462 bp in length
* 59071 59170: gap of 100 bp
* 59171 61789: contig of 2619 bp in length
* 61790 61889: gap of 100 bp
* 61890 72802: contig of 10913 bp in length
* 72803 72902: gap of 100 bp
* 72903 75383: contig of 2481 bp in length
* 75384 75483: gap of 100 bp
* 75484 78512: contig of 3029 bp in length
* 78513 78612: gap of 100 bp
* 78613 80836: contig of 2224 bp in length
* 80837 80936: gap of 100 bp
* 80937 88055: contig of 7119 bp in length
* 88056 88155: gap of 100 bp
* 88156 92223: contig of 4068 bp in length
* 92224 92323: gap of 100 bp
* 92324 96342: contig of 4219 bp in length
* 96343 9642: gap of 100 bp
* 96643 99878: contig of 3236 bp in length
* 99879 99978: gap of 100 bp
* 99979 105558: contig of 5580 bp in length
* 10559 105658: gap of 100 bp
* 105659 122337: contig of 16679 bp in length
* 122338 122437: gap of 100 bp
* 12438 128755: contig of 6318 bp in length
* 128756 128855: gap of 100 bp
* 128856 131775: contig of 2920 bp in length
* 131776 131875: gap of 100 bp
* 131876 136653: contig of 4778 bp in length
* 136654 136753: gap of 100 bp
* 136754 149239: contig of 12486 bp in length
* 149240 149339: gap of 100 bp
* 149340 156036: contig of 6697 bp in length
* 156037 156136: gap of 100 bp
* 156137 161800: contig of 5664 bp in length
* 161801 161900: gap of 100 bp
* 161901 168950: contig of 7050 bp in length
* 168951 169050: gap of 100 bp
* 169051 172376: contig of 3326 bp in length.

FEATURES
            source
            i. 172376
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="9"
                /clone="RP11-138018"
                /clone_lib="RPC1-11.1"
            i. 17327
                /note="assembly_fragment:00680
                clone_end:sp6
                vector_side:left"
            17428..20692
                /note="assembly_fragment:00043
                fragment_chain:1"
            20793..23652
                /note="assembly_fragment:00840
                fragment_chain:1"
            23753..25971
                /note="assembly_fragment:00060
                fragment_chain:2"
            26072..46738
                /note="assembly_fragment:00750
                fragment_chain:2"
            46839..56508
                /note="assembly_fragment:00077
                fragment_chain:3"
            56609..59070
                /note="assembly_fragment:01416
```

misc_feature
fragment_chain:3"
59171..61789
/note="assembly_fragment:00144
fragment_chain:4"
61890..72802
/note="assembly_fragment:00688
fragment_chain:4"
72903..75383
/note="assembly_fragment:00093"
75484..78512
/note="assembly_fragment:00116"
78613..80836
/note="assembly_fragment:00141"
80937..88055
/note="assembly_fragment:00244"
88156..92223
/note="assembly_fragment:00457"
92324..96342
/note="assembly_fragment:01057"
96643..99878
/note="assembly_fragment:01059"
99979..105558
/note="assembly_fragment:01199"
105859..122337
/note="assembly_fragment:01269"
122438..128755
/note="assembly_fragment:01299"
128856..131775
/note="assembly_fragment:01418"
131876..136653
/note="assembly_fragment:01559"
136754..149239
/note="assembly_fragment:01640"
149340..156036
/note="assembly_fragment:01302
fragment_chain:5"
156137..161800
/note="assembly_fragment:00235
fragment_chain:5"
161901..168950
/note="assembly_fragment:00728
fragment_chain:5"
169051..172376
/note="assembly_fragment:00685
fragment_chain:5
clone_end:T7
vector_side:right"
BASE COUNT 47584 a 38065 c 37064 g 47145 t 2518 others
ORIGIN

Query Match 1.3%; Score 74; DB 2: Length 172376;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5438 GCCTAATTTGTTATTTTAGTAGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCPCA 5497
|||||
Db 14043 GCCTAATTTGTTATTTTAGTAGAGTGGGTTTCACCATGTTGGCCAGGCTGCTCPCA 14102
|||||
QY 5498 AACTCCTGACCTCA 5511
|||||
Db 14103 AACTCCTGACCTCA 14116
|||||

RESULT 11
AL139113
LOCUS AL139113 193280 bp DNA linear PRI 12-MAY-2001
DEFINITION Human DNA sequence from clone RP11-176F3 on chromosome 9, complete
sequence.
ACCESSION AL139113
VERSION AL139113.21 GI:14041744
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193280)
Tracey.A.
Direct Submission
Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 15, 2001 this sequence version replaced gi:13872249.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C-elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-176F3 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-176F3 The true
left end of clone RP11-138018 is at 125710 in this sequence. The
true right end of clone RP11-133022 is at 92565 in this sequence.
FEATURES
Location/Qualifiers
1..193280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-176F3"
/clone_lib="RPCI-11.1"
359..530
/note="MER5B repeat: matches 16..178 of consensus"
545..689
/note="MER5A repeat: matches 34..182 of consensus"
1993..2148
/note="3 copies 52 mer 92% conserved"
4526..4577
/note="26 copies 2 mer aa 80% conserved"
4880..5119
/note="AluX repeat: matches 1..259 of consensus"
5145..5843
/note="L1PA16 repeat: matches 5460..6157 of consensus"
7725..7802
/note="MIR repeat: matches 64..150 of consensus"
7860..8316
/note="L1MD repeat: matches 1189..1689 of consensus"
8317..8618
/note="AluX repeat: matches 1..302 of consensus"
8619..8714
/note="L1MD repeat: matches 1097..1189 of consensus"
8724..9205
/note="L1MC1 repeat: matches 5852..6332 of consensus"
9213..9358
/note="FLAM_C repeat: matches 1..142 of consensus"
9380..9631
/note="AluJo repeat: matches 43..282 of consensus"
9632..9804
/note="L1MC1 repeat: matches 5649..5823 of consensus"

```
repeat_region 9802..9920
/note="LIMB2 repeat: matches 6005..6114 of consensus"
repeat_region 10339..10370
/note="16 copies 2 mer gt 100% conserved"
repeat_region 10494..10642
/note="MIR repeat: matches 84..254 of consensus"
repeat_region 10818..11127
/note="LIM4 repeat: matches 11..302 of consensus"
repeat_region 11140..11451
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 11635..11777
/note="LIM4 repeat: matches 470..609 of consensus"
repeat_region 12367..12533
/note="L2 repeat: matches 2362..2530 of consensus"
repeat_region 12662..12914
/note="L2 repeat: matches 1931..2203 of consensus"
repeat_region 12918..13288
/note="LIPAL6 repeat: matches 4675..5036 of consensus"
repeat_region 13289..13579
/note="AluSg repeat: matches 5..295 of consensus"
repeat_region 13580..14394
/note="LIPAL6 repeat: matches 5036..5867 of consensus"
repeat_region 14433..14644
/note="LIR12 repeat: matches 450..671 of consensus"
repeat_region 14647..14780
/note="AluSp/q repeat: matches 159..293 of consensus"
repeat_region 14782..15061
/note="AluSg repeat: matches 33..299 of consensus"
repeat_region 15064..16072
/note="L1 repeat: matches 3303..4289 of consensus"
repeat_region 16073..16371
/note="AluSx repeat: matches 1..299 of consensus"
repeat_region 16372..16513
/note="L1 repeat: matches 4289..4432 of consensus"
repeat_region 16525..16632
/note="AluSg/x repeat: matches 1..108 of consensus"
repeat_region 16721..17032
/note="AluSx repeat: matches 1..308 of consensus"
repeat_region 17247..17299
/note="L2 repeat: matches 2699..2750 of consensus"
repeat_region 17451..17521
/note="L2 repeat: matches 2624..2694 of consensus"
repeat_region 18163..18761
/note="LIMC3 repeat: matches 7110..7727 of consensus"
repeat_region 18766..19053
/note="AluJb repeat: matches 1..289 of consensus"
repeat_region 19115..19217
/note="MIR repeat: matches 132..248 of consensus"
repeat_region 19230..19404
/note="AluSg/x repeat: matches 134..308 of consensus"
misc_feature 20104..21605
/note="CpG Island"
/evidence=not_experimental
repeat_region 20420..20745
/note="163 copies 2 mer cc 57% conserved"
repeat_region 21669..21702
/note="17 copies 2 mer gg 85% conserved"
repeat_region 21833..21854
/note="11 copies 2 mer aa 100% conserved"
repeat_region 22944..23239
/note="AluSg repeat: matches 1..296 of consensus"
repeat_region 24478..24760
/note="AluJo repeat: matches 1..278 of consensus"
repeat_region 24924..25224
/note="AluSx repeat: matches 1..296 of consensus"
repeat_region 25400..25694
/note="AluSp repeat: matches 1..295 of consensus"
repeat_region 25730..26027
/note="AluJo repeat: matches 1..298 of consensus"
repeat_region 26192..26508
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 26509..26640
/note="FLAM_C repeat: matches 1..135 of consensus"

repeat_region 26684..26978
/note="AluSx repeat: matches 1..295 of consensus"
repeat_region 27513..27657
/note="AluSx repeat: matches 1..150 of consensus"
repeat_region 27658..27957
/note="AluY repeat: matches 1..297 of consensus"
repeat_region 27958..28125
/note="AluSx repeat: matches 150..302 of consensus"
repeat_region 28467..28505
/note="AluY repeat: matches 1..47 of consensus"
repeat_region 28548..28843
/note="AluSx repeat: matches 1..301 of consensus"
repeat_region 29177..29465
/note="AluSx repeat: matches 15..299 of consensus"
repeat_region 29720..30029
/note="AluY repeat: matches 1..296 of consensus"
repeat_region 30103..30417
/note="AluSx repeat: matches 5..297 of consensus"
repeat_region 30473..30639
/note="AluJb repeat: matches 129..291 of consensus"
repeat_region 30662..30954
/note="AluSg repeat: matches 1..293 of consensus"
repeat_region 30992..31302
/note="AluSx repeat: matches 1..311 of consensus"
repeat_region 31327..31620
/note="AluSx repeat: matches 1..298 of consensus"
repeat_region 31698..31856
/note="AluSg/x repeat: matches 131..298 of consensus"
repeat_region 31857..32040
/note="AluSx repeat: matches 137..308 of consensus"
repeat_region 32041..32337
/note="AluSg repeat: matches 1..296 of consensus"
repeat_region 32338..32470
/note="AluSx repeat: matches 1..137 of consensus"
repeat_region 32668..32848
/note="LIPB3 repeat: matches 5974..6146 of consensus"
repeat_region 33097..33399
/note="AluSp repeat: matches 1..304 of consensus"
repeat_region 33462..33759
/note="AluSx repeat: matches 1..299 of consensus"
repeat_region 33792..34088
/note="AluSx repeat: matches 5..301 of consensus"
repeat_region 34281..34377
/note="HX3 repeat: matches 1..97 of consensus"
repeat_region 34378..34421
/note="22 copies 2 mer aa 79% conserved"

Query Match 1.3%; Score 74; DB 9; Length 193280;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GGCTAATGTTGTTATTTTAGTAGACATGGGGTTTACCATGTTGGCAGGCTGGTCTCA 5497
Db 139758 GGCTAATGTTGTTATTTTAGTAGAGATGGGGTTTACCATGTTGGCAGGCTGGTCTCA 139817
QY 5498 AACTCCTGACCTCA 5511
Db 139818 AACTCCTGACCTCA 139831

RESULT 12
AC048334/c 215719 bp DNA linear PRI 04-DEC-2001
LOCUS Homo sapiens 3 BAC RP11-372C15 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
ACCESSION AC048334
VERSION AC048334.21 GI:15027693
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 215719)
```


AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaraturung, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Blimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, I.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 215719)
Worley, K.C.
Direct Submission
Submitted (14-APR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 215719)
Worley, K.C.
Direct Submission
Submitted (01-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 215719)
Worley, K.C.
Direct Submission
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 215719)
Worley, K.C.
Direct Submission
Submitted (04-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 28, 2001 this sequence version replaced gi:15022627.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	Source	Location/Qualifiers
misc_feature	1..2296	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3" /clone="RP11-572C15" /note="overlaps bases 3705..6000 of clone AC069516"
repeat_region	complement(447..556)	/function="clone overlap"
repeat_region	623..1065	/rpt_family="L2"
repeat_region	complement(1882..2179)	/rpt_family="L2"
repeat_region	3667..3843	/rpt_family="AluSq"
repeat_region	complement(3931..4070)	/rpt_family="LIME"
repeat_region	complement(4624..4710)	/rpt_family="MIR"
repeat_region	4824..5131	/rpt_family="MIR"
repeat_region	5248..5271	/rpt_family="AluSq"
repeat_region	complement(5459..5772)	/rpt_family="AT-rich"
repeat_region	6767..6846	/rpt_family="AluSq"
repeat_region	complement(6875..6937)	/rpt_family="MER33"
repeat_region	7090..7350	/rpt_family="MER34B"
repeat_region	7371..7529	/rpt_family="MER33"
repeat_region	complement(7818..8112)	/rpt_family="MIR"
repeat_region	complement(8126..8364)	/rpt_family="AluSq"
repeat_region	complement(8365..9004)	/rpt_family="LIME3"
repeat_region		/rpt_family="LIME3"
repeat_region		/rpt_family="LrR8"

repeat_region complement(9005. .9352)
repeat_region /rpt_family="LIME3"
repeat_region 9469. .9895
repeat_region /rpt_family="trigger2a"
repeat_region complement(9896. .10130)
repeat_region /rpt_family="LIME3"
repeat_region complement(10382. .10670)
repeat_region /rpt_family="Aluy"
repeat_region 10979. .11013
repeat_region /rpt_family="AT_rich"
repeat_region complement(12207. .12556)
repeat_region /rpt_family="MLT1A"
repeat_region 12687. .12708
repeat_region /rpt_family="AT_rich"
repeat_region complement(13613. .13665)
repeat_region /rpt_family="MIR"
repeat_region 13686. .13947
repeat_region /rpt_family="AluX"
repeat_region 13948. .13999
repeat_region /rpt_family="TAAAA)n"
repeat_region complement(14000. .14033)
repeat_region /rpt_family="MIR"
repeat_region 14301. .14326
repeat_region /rpt_family="AT_rich"
repeat_region 14327. .14441
repeat_region /rpt_family="MER97a"
repeat_region complement(14603. .14908)
repeat_region /rpt_family="Aluy"
repeat_region 16345. .16586
repeat_region /rpt_family="CT_rich"
repeat_region 17189. .17213
repeat_region /rpt_family="(T)n"
repeat_region 17824. .18277
repeat_region /rpt_family="MLT1C"
repeat_region 18471. .18529
repeat_region /rpt_family="MIR"
repeat_region 19405. .19513
repeat_region /rpt_family="MIR"
repeat_region 19609. .19845
repeat_region /rpt_family="L1M4"
repeat_region 19844. .21794
repeat_region /rpt_family="LIME1"
repeat_region complement(21795. .22127)

Query Match 1.3%; Score 74; DB 9; Length 215719;
Best Local Similarity 100.08; Pred. No. 1.5e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GCCTAATGTTGTATTTTAGTAGAGATGGGTTTACCATGTTGGCCAGGCTGCTCA 5497
|||||
Db 122366 GCCTAATGTTGTATTTTAGTAGAGATGGGTTTACCATGTTGGCCAGGCTGCTCA 122307
QY 5498 AACTCCTGACCTCA 5511
|||||
Db 122306 AACTCCTGACCTCA 122293

RESULT 13
AC022717/c AC022717 173943 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 8 clone RP11-139K6 map 8, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
ACCESSION AC022717
VERSION AC022717.4 GI:12583845
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 173943)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-139K6
JOURNAL Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 173943)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 173943)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Willson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Jan 27, 2001 this sequence version replaced gi:11181823.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5402

Center clone name: L39_K6

----- Summary Statistics

Sequencing vector: M13; M7815; 2% of reads

Sequencing vector: plasmid; n/a; 98% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 173297 bases at least Q40

Consensus quality: 173583 bases at least Q30

Consensus quality: 173649 bases at least Q20

Insert size: 169000; agarose-ef

Insert size: 173743; sum-of-contigs

Quality coverage: 9.4 in Q20 bases; agarose-ef

Quality coverage: 9.2 in Q20 bas.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 153468: contig of 153468 bp in length
* 153469 153568: gap of 100 bp
* 153569 153791: contig of 223 bp in length
* 153792 153891: gap of 100 bp
* 153892 173943: contig of 20052 bp in length.
Location/Qualifiers
1. .173943
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="8"
/map="8"
/clone="RP11-139K6"
/clone_lib="RPCI-11 Human Male BAC"
1. .153468
/note="assembly_fragment
clone_end:SP6
vector_side:left"
153569 .153791
/note="assembly_fragment"
153892 173943
/note="assembly_fragment
clone_end:T7
vector_side:right"
a 35630 c 36092 g 53491 t 202 others
BASE COUNT 48528 a 35630 c 36092 g 53491 t 202 others
ORIGIN
Query Match 1.3%; Score 73; DB 2; Length 173943;
Best Local Similarity 100.0%; Pred. NO. 5.3e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5439 GCTATGTTGTTATTTTAGTAGAGATGGGTTTCACCATGTGGCCAGGCTGGTCTCAA 5498
|||||
Db 149908 GCTATGTTGTTATTTTAGTAGAGATGGGTTTCACCATGTGGCCAGGCTGGTCTCAA 149849
|||||
QY 5499 ACTCTGACCTCA 5511
|||||
Db 149848 ACTCTGACCTCA 149836
.
.
RESULT 14
AC027595
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-359A12 map 8, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
ACCESSION AC027595
VERSION AC027595.3 GI:14626930
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 195661)
Birren.B., Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens chromosome 8, clone RP11-359A12
Unpublished
2 (bases 1 to 195661)
Birren.B., Linton.L., Nusbaum.C., Lander.E., Abraham.H., Allen.N.,
Anderson.S., Baldwin.J., Barna.N., Bastien.V., Bedalov.F.,
Boguslavsky.L., Boukhgalter.B., Brown.A., Burkett.G.,
Campolano.A., Castie.A., Choepel.Y., Collangelo.M., Collins.S.,
Collumore.A., Cooke.P., DeArellano.K., Dewar.K., Diaz.J.S.,
Dodge.S., Domino.M., Doyle.M., Ferreira.P., FitzHugh.W., Gage.D.,
Galagan.J., Gardyna.S., Ginde.S., Goyette.M., Graham.L.,
Grand-Pierre.N., Grant.G., Hagos.B., Heaford.A., Horton.L.,
Howland.J.C., Iliev.I., Johnson.R., Jones.C., Kann.L., Karatas.A.,
Klein.J., Lacroque.K., Lamazares.R., Landers.T., Lehoczy.J.,
Levine.R., Lieu.C., Liu.G., Locke.K., Macdonald.P., Marquis.N.,
McCarthy.M., McEwan.P., McGurk.A., McKernan.K., McPheeters.R.,

Meldrim.J., Meneus.L., Mihova.T., Miranda.C., Mlenga.V., Morrow.J.,
Murphy.T., Naylor.J., Norman.C.H., O'Connor.T., O'Donnell.P.,
O'Neil.D., Oliver.T.M., Oliver.J., Peterson.K., Pierre.N.,
Pisani.C., Pollara.V., Raymond.C., Riley.R., Rogov.P., Rothman.D.,
Roy.A., Santos.R., Schauer.S., Severy.P., Spencer.B.,
Stange-Thomann.N., Stojanovic.N., Subramanian.A., Talamas.J.,
Testaye.S., Theodore.J., Tirrell.A., Travers.M., Trigilio.J.,
Vassiliev.H., Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J.,
Young.G., Zainoun.J., Zimmer.A. and Zody.M.
Direct Submission
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 8, 2001 this sequence version replaced gi:7677926.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9099
Center clone name: 359_A_12
----- Summary Statistics
Sequencing vector: M13; M77815; 35% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191730 bases at least Q40
Consensus quality: 193959 bases at least Q30
Consensus quality: 194605 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 194961; sum-of-contigs
Quality coverage: 10.5 in Q20 bases; agarose-fp
Quality coverage: 9.3 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 36680: contig of 36680 bp in length
* 36681 36780: gap of 100 bp
* 36781 38091: contig of 1311 bp in length
* 38092 38191: gap of 100 bp
* 38192 39416: contig of 1225 bp in length
* 39417 39516: gap of 100 bp
* 39517 40566: contig of 1050 bp in length
* 40567 40666: gap of 100 bp
* 40667 42013: contig of 1347 bp in length
* 42014 42113: gap of 100 bp
* 42114 43651: contig of 1538 bp in length
* 43652 43751: gap of 100 bp
* 43752 44775: contig of 1024 bp in length
* 44776 44875: gap of 100 bp
* 44876 195661: contig of 150786 bp in length.
Location/Qualifiers
1. .195661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="8"
/map="8"
/clone="RP11-359A12"
/clone_lib="RPCI-11 Human Male BAC"
1. .36680
/note="assembly_fragment
clone_end:SP6
vector_side:left"
36781 .38091
/note="assembly_fragment"
FEATURES
source
misc_feature
misc_feature

```
misc_feature 38192..39416
              /note="assembly_fragment"
misc_feature 39517..40566
              /note="assembly_fragment"
misc_feature 40667..42013
              /note="assembly_fragment"
misc_feature 42114..43651
              /note="assembly_fragment"
misc_feature 43752..44775
              /note="assembly_fragment"
misc_feature 44876..495661
              /note="assembly_fragment"
              clone_end:r7
vector_side:right"
BASE COUNT 59915 a 40215 c 39680 g 55151 t 700 others
ORIGIN
-----
Query Match 1.3%; Score 73; DB 2; Length 195661;
Best Local Similarity 100.0%; Pred. No. 5.3e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5439 GCTAATGTTTCTATTATTTAGTAGAGATGGGCTTTCACCATGTTGCCAGGCTGCTCAAA 5498
      |||||||
Db 51109 GCTAATGTTTCTATTATTTAGTAGAGATGGGCTTTCACCATGTTGCCAGGCTGCTCAAA 51168

Qy 5499 ACTCCTGACCTCA 5511
      |||||||
Db 51169 ACTCCTGACCTCA 51181

RESULT 15
LOCUS AF235101 214696 bp DNA linear HTG 26-JUL-2002
DEFINITION Homo sapiens chromosome 8 clone CTD-2571K21 map 8q, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
ACCESSION AF235101
VERSION AF235101.4 Gi:16356856
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          Wen,G., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
          Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schilhabel,M.B.,
          Schudy,A., Siddiqui,R., Taudien,S., Rosenthal,A. and Platzer,M.
          Chromosome 8 genomic sequence
          Unpublished
          2 (bases 1 to 214696)
          Blechschmidt,K., Wen,G., Schilhabel,M., Baumgart,C., Menzel,U.,
          Dette,M., Jahn,N. and Rosenthal,A.
          Direct Submission
          Submitted (16-FEB-2000) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
          On Oct 24, 2001 this sequence version replaced gi:14329032.
          -----
          Center: Institute of Molecular Biotechnology
          Center code: IMB
          Web site: http://genome.imb-jena.de/
          Contact: gscj-submit@genome.imb-jena.de
          -----
          Project Information
          Center project name: H360
          Center clone name: CTD-2571K21
          -----
          Summary Statistics
          Sequencing vector: M13; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 200056 bases at least Q40
          Consensus quality: 206842 bases at least Q30
          Consensus quality: 210985 bases at least Q20
          Quality coverage: 4.93 x in Q20 bases; sum-of-contigs
          -----
          Sequence Quality Assessment:
```

```
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality 10.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 14268: contig of 14268 bp in length
* 14269 14368: gap of unknown length
* 14369 16329: contig of 1961 bp in length
* 16330 16429: gap of unknown length
* 16430 21020: contig of 4591 bp in length
* 21021 21120: gap of unknown length
* 21121 31366: contig of 10246 bp in length
* 31367 31466: gap of unknown length
* 31467 38824: contig of 7358 bp in length
* 38825 38925: gap of unknown length
* 38926 48608: contig of 9684 bp in length
* 48609 48708: gap of unknown length
* 48709 67136: contig of 18428 bp in length
* 67137 67237: gap of unknown length
* 67238 138505: contig of 71269 bp in length
* 138506 138605: gap of unknown length
* 138606 168996: contig of 28291 bp in length
* 168997 168998: gap of unknown length
* 168999 197031: contig of 30035 bp in length
* 197032 197131: gap of unknown length
* 197132 214696: contig of 17565 bp in length.
          Location/Qualifiers
          1..214696
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="8"
            /map="8q"
            /clone="CTD-2571K21"
BASE COUNT 63313 a 42987 c 42987 g 64499 t 1000 others
ORIGIN
-----
Query Match 1.3%; Score 73; DB 2; Length 214696;
Best Local Similarity 100.0%; Pred. No. 5.3e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5439 GCTAATGTTTCTATTATTTAGTAGAGATGGGCTTTCACCATGTTGCCAGGCTGCTCAAA 5498
      |||||||
Db 172295 GCTAATGTTTCTATTATTTAGTAGAGATGGGCTTTCACCATGTTGCCAGGCTGCTCAAA 172354

Qy 5499 ACTCCTGACCTCA 5511
      |||||||
Db 172355 ACTCCTGACCTCA 172367

Search completed: June 19, 2003, 17:42:00
Job time : 13765 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:37:34 ; Search time 1071 Seconds
(without alignments)
11594.322 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacataggtcca.....tcaaacctctgacctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5514	100.0	5514	21 AAC55313	Human activation-i
2	4771	86.5	11204	21 AAC55339	Human activation-i
c 3	118	2.1	772	23 AAS81193	DNA encoding novel
4	87	1.6	87	21 AAC55315	Human activation-i
5	87	1.6	2818	21 AAC55312	Human activation-i
c 6	71	1.3	12718	24 ABK52229	CDNA encoding huma
c 7	69	1.3	539	23 ABV57689	Human prostate exp
c 8	67	1.2	84	22 AAL07192	Human reproductive
9	67	1.2	95	22 ABA18401	Human nervous syst

c 10	67	1.2	121	22 ABA14482	Human nervous syst
c 11	67	1.2	132	22 AAK80307	Human immune/haema
c 12	67	1.2	141	22 AAK66840	Human immune/haema
c 13	67	1.2	141	22 AAK68098	Human immune/haema
c 14	67	1.2	141	22 AAK68878	Human immune/haema
c 15	67	1.2	141	22 AAK70059	Human immune/haema
c 16	67	1.2	141	22 AAK72514	Human immune/haema
c 17	67	1.2	141	22 AAK72808	Human immune/haema
c 18	67	1.2	141	22 AAK72892	Human immune/haema
c 19	67	1.2	141	22 AAK80869	Human immune/haema
c 20	67	1.2	141	22 AAK83523	Human immune/haema
c 21	67	1.2	141	22 AAK85680	Human immune/haema
c 22	67	1.2	161	22 AAK66827	Human immune/haema
c 23	67	1.2	481	22 ABA18403	Human nervous syst
c 24	67	1.2	566	22 ABA10398	Human cDNA clone (
c 25	67	1.2	606	24 AB057438	Human colon cancer
c 26	67	1.2	1160	22 AAL02996	Human reproductive
c 27	67	1.2	1160	23 ABL97343	Human testicular a
c 28	67	1.2	1607	22 AAK17521	Human cDNA sequenc
c 29	67	1.2	3015	22 AAK85238	Human immune/haema
c 30	67	1.2	3780	22 AAK72236	Human immune/haema
c 31	67	1.2	3780	22 AAK79330	Human immune/haema
c 32	67	1.2	4736	21 AAC69133	Human ABC1 gene ex
c 33	67	1.2	5930	22 AAK36019	Human musculoskele
c 34	67	1.2	7232	22 AAK72237	Human immune/haema
c 35	67	1.2	7232	22 AAK79332	Human immune/haema
c 36	67	1.2	9139	22 AAK80305	Human immune/haema
c 37	67	1.2	9139	22 AAK80306	Human immune/haema
c 38	67	1.2	9723	22 AAK65237	Human immune/haema
c 39	67	1.2	11426	22 AAK84781	Human immune/haema
c 40	67	1.2	11754	22 AAS04033	Human ABC1 gene, p
c 41	67	1.2	16086	22 AAL36020	Human musculoskele
c 42	67	1.2	16086	22 AAK69721	Human immune/haema
c 43	67	1.2	16869	21 ABN97976	Human retroviral s
c 44	67	1.2	20303	18 AAT71699	Human deoxycytidyl
c 45	67	1.2	23580	22 AAK66230	Human immune/haema

ALIGNMENTS

RESULT 1

AAC55313

ID AAC55313 standard; DNA; 5514 BP.

XX AAC55313;

AC AAC55313;

XX 05-FEB-2001 (first entry)

DT Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:9.

DE Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW immune related disease; allergy; allergic disease; antiallergic;
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

XX Homo sapiens.

Db	3781	TGCTTGGCCACCTGCACCTCAGCCTGGGTGACAGGACCAGACCTTGCCCTCAAAAAATAA	3840
QY	3841	GAAGAAAAATATAAATAAATGAACAACACTACAAAGAGCTGTGTCTAGATGAGCTAC	3900
Db	3841	GAAGAAAAATATAAATAAATGAACAACACTACAAAGAGCTGTGTCTAGATGAGCTAC	3900
QY	3901	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGCTGTCTACCTGCACCTAC	3960
Db	3901	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGCTGTCTACCTGCACCTAC	3960
QY	3961	ATTATTAAATAATCAATCTCAATGTATATCCACACAAGACATGGTACGTGAATGTTTCAT	4020
Db	3961	ATTATTAAATAATCAATCTCAATGTATATCCACACAAGACATGGTACGTGAATGTTTCAT	4020
QY	4021	AGTACCTTTATTCCAAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080
Db	4021	AGTACCTTTATTCCAAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080
QY	4081	CAATTAACAATAATGTCTATATCCATGCAATGGAATACACCCCTGCAGTACAAAGGAAG	4140
Db	4081	CAATTAACAATAATGTCTATATCCATGCAATGGAATACACCCCTGCAGTACAAAGGAAG	4140
QY	4141	AACTACTTTGGGGATGAATCCCAAGTCAATGACGCTAAATGAAGAGTCAGACATGAAGG	4200
Db	4141	AACTACTTTGGGGATGAATCCCAAGTCAATGACGCTAAATGAAGAGTCAGACATGAAGG	4200
QY	4201	AGGAGATAATGTATGCCATACGAAATTTCTAGAAAAATGAAAGTAACTTATAGTTACAGAAA	4260
Db	4201	AGGAGATAATGTATGCCATACGAAATTTCTAGAAAAATGAAAGTAACTTATAGTTACAGAAA	4260
QY	4261	GCAAAATCAGGGCAGGATAGAGCTCACACCTGTAATCCACAGACATTTGAGAGGCCACGT	4320
Db	4261	GCAAAATCAGGGCAGGATAGAGCTCACACCTGTAATCCACAGACATTTGAGAGGCCACGT	4320
QY	4321	GGGAAGATGCTAGAACTCAGGAGTTCAAGACCAGCTTGGGCAACACAGTGAACCTCCAT	4380
Db	4321	GGGAAGATGCTAGAACTCAGGAGTTCAAGACCAGCTTGGGCAACACAGTGAACCTCCAT	4380
QY	4381	TTCCCAAAAAATGGGAAAAAAGAAAGCAAAATCAGTGTGTGCTGTGGGAGGGGAAG	4440
Db	4381	TTCCCAAAAAATGGGAAAAAAGAAAGCAAAATCAGTGTGTGCTGTGGGAGGGGAAG	4440
QY	4441	GACTGCAAGAGGGGAAGAGCTCTGTGGGGTGAGGGTGGTATTCAGGTTCTGTATCCT	4500
Db	4441	GACTGCAAGAGGGGAAGAGAGCTCTGTGGGGTGAGGGTGGTATTCAGGTTCTGTATCCT	4500
QY	4501	GACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGTCATCTTA	4560
Db	4501	GACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTATGTCATCTTA	4560
QY	4561	AATGGGTGAGTTTACTGTATGTAAATATACCTCAATGAAGAAAAATAATGTGTAAG	4620
Db	4561	AATGGGTGAGTTTACTGTATGTAAATATACCTCAATGAAGAAAAATAATGTGTAAG	4620
QY	4621	AAAAGTTTCAATCTCTCCAGCAACCTTATTCAAATCTTGAGCCCTTTACTTCGCA	4680
Db	4621	AAAAGTTTCAATCTCTCTCCAGCAACCTTATTCAAATCTTGAGCCCTTTACTTCGCA	4680
QY	4681	AATTTCTCTGCACTTCTGCCCTTACCATTTAGGTGACAGCACTAGCTTCCACAAATTCGATA	4740
Db	4681	AATTTCTCTGCACTTCTGCCCTTACCATTTAGGTGACAGCACTAGCTTCCACAAATTCGATA	4740
QY	4741	AATGCATTTCTGGAAGAAGACTAGGGACAAAAATCCAGGCATCACTTGTGCTTTTCATATCA	4800
Db	4741	AATGCATTTCTGGAAGAAGACTAGGGACAAAAATCCAGGCATCACTTGTGCTTTTCATATCA	4800
QY	4801	CCAGCTGTACAGCTGTGCTGCTCTGCACCTGCAATGGGACTCTTGATTTCTTTAA	4860
Db	4801	CCAGCTGTACAGCTGTGCTGCTCTGCACCTGCAATGGGACTCTTGATTTCTTTAA	4860
QY	4861	GGAAACTTGGGTACCAGAGTATTTCCAAAAATGCTATTCAAAATAGTCTTATGATATG	4920
Db	4861	GGAAACTTGGGTACCAGAGTATTTCCAAAAATGCTATTCAAAATAGTCTTATGATATG	4920
QY	4921	CAAGACACTGTCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCAATTATGTGGGA	4980
Db	4921	CAAGACACTGTCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCAATTATGTGGGA	4980
QY	4981	ACAACATAGCAAGATATTTAGATCAATTTTGACTAGTTTAAAAAAAGCAGCAGTACAAAAT	5040
Db	4981	ACAACATAGCAAGATATTTAGATCAATTTTGACTAGTTTAAAAAAAGCAGCAGTACAAAAT	5040
QY	5041	CACACATGCANTCAGTATATAATCCAAATCATGTAAATATGTCCTGTAGAAAGACTAGAGG	5100
Db	5041	CACACATGCANTCAGTATATAATCCAAATCATGTAAATATGTCCTGTAGAAAGACTAGAGG	5100
QY	5101	AATAAACAACAAGAAATCTTAACAGCTCATTTGACCTATTAGACACTAAGTCTAATTATTATT	5160
Db	5101	AATAAACAACAAGAAATCTTAACAGCTCATTTGACCTATTAGACACTAAGTCTAATTATTATT	5160
QY	5161	AGACACTATGATATTTGAGATTTTAAAAAATCTTTTAATATTTTAAAAATTTAGAGCTCTTCT	5220
Db	5161	AGACACTATGATATTTGAGATTTTAAAAAATCTTTTAATATTTTAAAAATTTAGAGCTCTTCT	5220
QY	5221	ATTTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTT	5280
Db	5221	ATTTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTT	5280
QY	5281	TTTTTTTTTTTTTGGATGGAGTTTGGTCTTGTGCCATGCTGGAGTGAATGGCATG	5340
Db	5281	TTTTTTTTTTTTTGGATGGAGTTTGGTCTTGTGCCATGCTGGAGTGAATGGCATG	5340
QY	5341	AYCATAGCTCAGTCGCAACCTCCACCTCTCGGGTTCGAAGCAAAAGCTGTGCGCTCAGCCTCC	5400
Db	5341	AYCATAGCTCAGTCGCAACCTCCACCTCTCGGGTTCGAAGCAAAAGCTGTGCGCTCAGCCTCC	5400
QY	5401	CGGGTAGATGGGATTAACAGCCGCCACACACACTCGGCTAATGTTGTATTTTATGTA	5460
Db	5401	CGGGTAGATGGGATTAACAGCCGCCACACACACTCGGCTAATGTTGTATTTTATGTA	5460
QY	5461	GAGATGGGTTCACCATGTTGGCCAGGCTGTCTCAAACTCCTGACCTCAGAG	5514
Db	5461	GAGATGGGTTCACCATGTTGGCCAGGCTGTCTCAAACTCCTGACCTCAGAG	5514
RESULT 2			
AAC55339			
ID	AAC55339 standard; DNA; 11204 BP.		
XX	AAC55339;		
AC			
XX			
DT	05-FEB-2001 (first entry)		
XX			
DE	Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.		
XX			
DE	Activation-induced cytidine deaminase; AID; cytidine deaminase;		
XX			
KW	Immune related disease; allergy; allergic disease; antiallergic;		
KW	antianemic; antisthmatic; ophthalmological; anti-HIV; dermatological;		
KW	gene therapy; B cell associated immune system disorder; food allergy;		
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;		
KW	IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;		
KW	drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;		
KW	ataxia telangiectasia; common variable immunodeficiency disorder;		
KW	major histocompatibility class II deficiency disease;		
KW	auto immunodeficiency syndrome; IgG subclass selection disorder; ds.		
OS	Homo sapiens.		
XX			
XX	WO200058480-A1.		
PN			
XX			
PD	05-OCT-2000.		
XX			
PF	28-MAR-2000; 2000WO-JP01918.		
XX			
PR	29-MAR-1999; 99JP-0087192.		
PR	24-JUN-1999; 99JP-0178999.		

PR 27-DEC-1999; 99JP-0371382.
 XX (NIBS) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX Honjo T, Muramatsu M;
 PI WPI; 2000-611715/58.
 DR
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 PS Claim 17; Page 163-170; 174pp; Japanese.
 XX
 CC The present invention describes an activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, Digorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders. The
 CC present sequence represents a genomic DNA sequence of human AID.
 XX
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;

Query Match 86.5%; Score 4771; DB 21; Length 11204;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 4921; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 591 AGGTTACAGAGACTGGGAATATGGGGAAATAGAGGCTATCTGAGGCTCTTCAACAC 650
 DB 1 AGGTTACAGAGACTGGGAATATGGGGAAATAGAGGCTATCTGAGGCTCTTCAACAC 60
 QY 651 AATAACCAAGAAGCTATTAATGCTCTTTAAGGTATTTACATAAATATTACTATTCTC 710
 DB 61 AATAACCAAGAAGCTATTAATGCTCTTTAAGGTATTTACATAAATATTACTATTCTC 120
 QY 711 ATTGTGCTTTATTTTGTGTTATCATGATTATTAATGAAGTGCTACTGTTACTGCCCTCC 770
 DB 121 ATTGTGCTTTATTTTGTGTTATCATGATTATTAATGAAGTGCTACTGTTACTGCCCTCC 180
 QY 771 TGATCTTTGTAGCTATGAGCATGGAGCTGGGCTTTTATAGCAGCAGCCCCAAGGAACC 830
 DB 181 TGATCTTTGTAGCTATGAGCATGGAGCTGGGCTTTTATAGCAGCAGCCCCAAGGAACC 240
 QY 831 TAAACATTAAGCAGAGCTGCCCTCAATGGTTTAACTGTTGACTCTGCTATGACAGC 890
 DB 241 TAAACATTAAGCAGAGCTGCCCTCAATGGTTTAACTGTTGACTCTGCTATGACAGC 300
 QY 891 CCCACCCACCCTTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGTTACCTGTGG 950
 DB 301 CCCACCCACCCTTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGTTACCTGTGG 360
 QY 951 GGGTGATGCTCTCAGGGAGAGGCCCAAAAGGGCAAGCTCAAAATTGAATGTGAAGGCC 1010
 DB 361 GGGTGATGCTCTCAGGGAGAGGCCCAAAAGGGCAAGCTCAAAATTGAATGTGAAGGCC 420
 QY 1011 AATGCACTGTGAGACTGAGACAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCT 1070
 DB 421 AATGCACTGTGAGACTGAGACAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCT 480
 QY 1071 GAGACTTGCAGGGAGGCAAGAGACACTCTCGACACCCTATGGACAGGTAAGAGGCCAG 1130
 DB

DB 481 GAGACTTGCAGGGAGGCAAGAGACACTCTGGACACCACACTATGGACAGGTAAAGAGCGAG 540
 QY 1131 TCTTCTCGTGGGTGATTCGACTGGCTTCTCTCAGAGCAAAATCTGAGTAATGACACTGG 1190
 DB 541 TCTTCTCGTGGGTGATTCGACTGGCTTCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 600
 QY 1191 TAGCTATCCCTTTCTCTCATGTACTGACTGCTGACTGATAAGATCACTTGAATCAATATGCA 1250
 DB 601 TAGCTATCCCTTTCTCTCATGTACTGCTGACTGATAAGATCACTTGAATCAATATGCA 660
 QY 1251 TATATATTTTGTGATCTCTCTCTTTCTTATTCAGATCTTATACGCTGTCAGCCCAA 1310
 DB 661 TATATATTTTGTGATCTCTCTCTTTCTTATTCAGATCTTATACGCTGTCAGCCCAA 720
 QY 1311 TTCCTTCGTTTCAGACTCTCTTGATTTCCCTCTCTTTTCATGTGGCAAAAGACTAGTG 1370
 DB 721 TTCCTTCGTTTCAGACTCTCTTGATTTCCCTCTCTTTTCATGTGGCAAAAGACTAGTG 780
 QY 1371 CGTACAATGTACTGATTCGCTCTGAGATTTGTACCATGGTTGAAACTAAATTTATGGTAAT 1430
 DB 781 CGTACAATGTACTGATTCGCTCTGAGATTTGTACCATGGTTGAAACTAAATTTATGGTAAT 840
 QY 1431 AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAGGTAATAGCAGTACTGTA 1490
 DB 841 AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAGGTAATAGCAGTACTGTA 900
 QY 1491 CTAAGGCGTAGTCTAATTTTCGTAATATTTTGTAAATTTCAACAGTAAACAACT 1550
 DB 901 CTAAGGCGTAGTCTAATTTTCGTAATATTTTGTAAATTTCAACAGTAAACAACT 960
 QY 1551 TGAAGACACACTTCTCTAGGGAGGCTTACTGAAATAATTTAGCTATATAGTAAGAAATTT 1610
 DB 961 TGAAGACACACTTCTCTAGGGAGGCTTACTGAAATAATTTAGCTATATAGTAAGAAATTT 1020
 QY 1611 GTAATTTTAGAAATGCCAAGCATTTCTAATTAATTTGCTTGAAGTCACTATGATGTGTC 1670
 DB 1021 GTAATTTTAGAAATGCCAAGCATTTCTAATTAATTTGCTTGAAGTCACTATGATGTGTC 1080
 QY 1671 CATATATAGGAGACAAATTCATCAAGCAAGTATTTAATGTTAAAGGCCCAATTTGTTAG 1730
 DB 1081 CATATATAGGAGACAAATTCATCAAGCAAGTATTTAATGTTAAAGGCCCAATTTGTTAG 1140
 QY 1731 GCAGTTAATGGCAGCTTTTACTATTAACATACTTTTCCATTTTGTTCAGACGTAGCTTAAT 1790
 DB 1141 GCAGTTAATGGCAGCTTTTACTATTAACATACTTTTCCATTTTGTTCAGACGTAGCTTAAT 1200
 QY 1791 TACCTCTTAGGTGGAATTTGGTTAAGGCTCCTCATTAATGCTTTTATGTCAGCTTTTGTAT 1850
 DB 1201 TACCTCTTAGGTGGAATTTGGTTAAGGCTCCTCATTAATGCTTTTATGTCAGCTTTTGTAT 1260
 QY 1851 AGGTTATGTCATAGAACTTATCTATTCCTACATTTTATGATTAATGATGATGATGAGA 1910
 DB 1261 AGGTTATGTCATAGAACTTATCTATTCCTACATTTTATGATTAATGATGATGATGAGA 1320
 QY 1911 ATAACACCTTAATCTCTTACTTACTCTCAATTTTAACTCTTTTATTAAGAACTTACATTTAC 1970
 DB 1321 ATAACACCTTAATCTCTTACTTACTCTCAATTTTAACTCTTTTATTAAGAACTTACATTTAC 1380
 QY 1971 AGAATAAAGATTTTAAAAATATATTTTTTTTGTAGAGACAGGGGCTTTAGCCCAAGCGAG 2030
 DB 1381 AGAATAAAGATTTTAAAAATATATTTTTTTTGTAGAGACAGGGGCTTTAGCCCAAGCGAG 1440
 QY 2031 GCTGGTCTCTAAGTCTTGGCCCAAGCGATCTCTCTGCTGGGCTTCTTAAAGTGTGAGNA 2090
 DB 1441 GCTGGTCTCTAAGTCTTGGCCCAAGCGATCTCTCTGCTGGGCTTCTTAAAGTGTGAGNA 1500
 QY 2091 TTATAGACATGAGCCATCACATCAATATACAGAAATAAGATTTTAAATGAGGAGTTTAA 2150
 DB 1501 TTATAGACATGAGCCATCACATCAATATACAGAAATAAGATTTTAAATGAGGAGTTTAA 1560
 QY 2151 TGTCTTTCAGAAAAATTTTCTTTGAGGTGAGCAAAATGTCAAATGTCCTCCTCAGTTTACCTG 2210
 DB 1561 TGTCTTTCAGAAAAATTTTCTTTGAGGTGAGCAAAATGTCAAATGTCCTCCTCAGTTTACCTG 1620

QY	2211	AGATTTTGAACAAGCTCTGAGCTATAGGTTCCTTGTGAAGGTCCTCATTTGGAATACTTGT	2270	3291	TAGCTTTTCTCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCTTGGCCTCCCAA	3350
Db	1621	AGATTTTGAACAAGCTCTGAGCTATAGGTTCCTTGTGAAGGTCCTCATTTGGAATACTTGT	1680	2701	TAGCTTTTCTGAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCTTGGCCTCCCAA	2760
QY	2271	TCAAAGTAAATGGAAGCAAGGTAAAATACGACGTTGAAATTCAGAGAAAGACAGAAA	2330	3351	GTGTTGGGATACAGTCAATGAGGCATCGCATCGGCCTAGGATCCATTTAGATTTAAATAT	3410
Db	1681	TCAAAGTAAATGGAAGCAAGGTAAAATACGACGTTGAAATTCAGAGAAAGACAGAAA	1740	2761	GTGTTGGGATACAGTCAATGAGGCATCGCATCGGCCTAGGATCCATTTAGATTTAAATAT	2820
QY	2331	AGGAGAAAAGATGAAATTCACAGGACAGAGGGAATATATATCATTAAGGAGCAG	2390	3411	GCATTTTAAATTTTAAATAAATATGCTAAATTTTACCTTATGTATGTATCTACTGGTA	3470
Db	1741	AGGAGAAAAGATGAAATTCACAGGACAGAGGGAATATATATCATTAAGGAGCAG	1800	2821	GCATTTTAAATTTTAAATAAATATGCTAAATTTTACCTTATGTATGTATCTACTGGTA	2880
QY	2391	TATCTGTAGAGCTCATTTAGTGTATGGCAAAATGACCTTGGTCAGGATTAATTTTAAACCGCT	2450	3471	ATAAATCTAGTTTGTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTTCATGTACGTGAG	3530
Db	1801	TATCTGTAGAGCTCATTTAGTGTATGGCAAAATGACCTTGGTCAGGATTAATTTTAAACCGCT	1860	2881	ATAAATCTAGTTTGTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTTCATGTACGTGAG	2940
QY	2451	TGTTTCTGTTTGCACGGCTGGGATGACGCTAGGTTTCTGCTCAGGAGCAGACGCTGT	2510	3531	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTCCACGCCCTGTAATCCCAGC	3590
Db	1861	TGTTTCTGTTTGCACGGCTGGGATGACGCTAGGTTTCTGCTCAGGAGCAGACGCTGT	1920	2941	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTCCACGCCCTGTAATCCCAGC	3000
QY	2511	CCAGACAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTCCTACTCA	2570	3591	ACTCTGGGAGGCTGAGGTGGTGGATCGCTTTGAGCCCTCGAGTTTCAAGACACAGCCTGAGC	3650
Db	1921	CCAGACAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTCCTACTCA	1980	3001	ACTCTGGGAGGCTGAGGTGGTGGATCGCTTTGAGCCCTCGAGTTTCAAGACACAGCCTGAGC	3060
QY	2571	GGACAGAAATGACGAGAACAGGAGCTGGAACAGGCCCTTAACACAGAGAGGGAAGTAA	2630	3651	AACATGGCAAAACCCCTGTTTCTATAACAAAAATTAGCCGGGCATGCTGCATGTGCTGT	3710
Db	1981	GGACAGAAATGACGAGAACAGGAGCTGGAACAGGCCCTTAACACAGAGAGGGAAGTAA	2040	3061	AACATGGCAAAACCCCTGTTTCTATAACAAAAATTAGCCGGGCATGCTGCATGTGCTGT	3120
QY	2631	TGGATCAACAAAGTTAACTAGCAGGTGAGGATCAGCAATTCATTTCACTCTGACTGGTA	2690	3711	GGTCCCAGCTACTAGGGGCTGAGGAGGAGAAATCTTTGGAGCCCAGGAGGTCAAGGCTG	3770
Db	2041	TGGATCAACAAAGTTAACTAGCAGGTGAGGATCAGCAATTCATTTCACTCTGACTGGTA	2100	3121	GGTCCCAGCTACTAGGGGCTGAGGAGGAGAAATCTTTGGAGCCCAGGAGGTCAAGGCTG	3180
QY	2691	ACATGTGACAGAAACAGTGTAGCTTATTTGTTATTTTCATGTAGAGTAGGACCCAAAAATC	2750	3771	CACGTGAGCAGTGTTCGCGCACTGCACCTCCAGCCTGGGTGACAGGACCAGACCTTCGCCTC	3830
Db	2101	ACATGTGACAGAAACAGTGTAGCTTATTTGTTATTTTCATGTAGAGTAGGACCCAAAAATC	2160	3181	CACGTGAGCAGTGTTCGCGCACTGCACCTCCAGCCTGGGTGACAGGACCAGACCTTCGCCTC	3240
QY	2751	CACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATATCTTCCAGGACACTTTTTC	2810	3831	AAAAAATAAGAGAAAAATTAATAATAAATGGAACAACTACAAAGAGCTGTGTGCTTA	3890
Db	2161	CACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATATCTTCCAGGACACTTTTTC	2220	3241	AAAAAATAAGAGAAAAATTAATAATAAATGGAACAACTACAAAGAGCTGTGTGCTTA	3300
QY	2811	TTCCTTATGATAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2870	3891	GATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTCTGTCA	3950
Db	2221	TTCCTTATGATAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2280	3301	GATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTCTGTCA	3360
QY	2871	CACACACACACAAACACACACACCCGCCCAACCAAGGTGCATGTAATAAGAGTGTAGATT	2930	3951	CCTGCACTACATTTTAAAAATATCAATTTCTCAATGTATATCCACACAAAGACTTGGTACGT	4010
Db	2281	CACACACACACAAACACACACACCCGCCCAACCAAGGTGCATGTAATAAGAGTGTAGATT	2340	3361	CCTGCACTACATTTTAAAAATATCAATTTCTCAATGTATATCCACACAAAGACTTGGTACGT	3420
QY	2931	CTCTGCTTTCTCATCTACACGCCCAGGAGGTAAAGTTAATAAGAGGGATTTATTGG	2990	4011	GAATGTTCACTAGTACCTTTTATTCACAAAACCCCAAGTATCCAAATATCCATC	4070
Db	2341	CTCTGCTTTCTCATCTACACGCCCAGGAGGTAAAGTTAATAAGAGGGATTTATTGG	2400	3421	GAATGTTCACTAGTACCTTTTATTCACAAAACCCCAAGTATCCAAATATCCATC	3480
QY	2991	TAAAGATGATGCTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACT	3050	4071	AACAAGTGAACAAATAACAAAAATGCTATATCCATGCANTGGAATACCAACCTCCAGT	4130
Db	2401	TAAAGATGATGCTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACTGTTTAACT	2460	3481	AACAAGTGAACAAATAACAAAAATGCTATATCCATGCANTGGAATACCAACCTCCAGT	3540
QY	3051	GTACTTATTAAGCACCTTATATGTTGAGCTTATATATACAAAGGTTTATATATGCTTA	3110	4131	ACAAAGGAAGAGCTACTTTGGGATCAATCCCAAGTCAATGAGCTTAAATGAAAGAGTCA	4190
Db	2461	GTACTTATTAAGCACCTTATATGTTGAGCTTATATATACAAAGGTTTATATATGCTTA	2520	3541	ACAAAGGAAGAGCTACTTTGGGATCAATCCCAAGTCAATGAGCTTAAATGAAAGAGTCA	3600
QY	3111	ATATAGTAATAGTAATGKTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	3170	4191	GACATGAAGGAGGAGATATATGATGCCATACGAAATTTCTAGAAAAATGAAAGTAACTTATA	4250
Db	2521	ATATAGTAATAGTAATGKTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	2580	3601	GACATGAAGGAGGAGATATATGATGCCATACGAAATTTCTAGAAAAATGAAAGTAACTTATA	3660
QY	3171	TAAATAAGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3230	4251	GTTTACAGAAAGCAAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCAGACCTTTGA	4310
Db	2581	TAAATAAGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2640	3661	GTTTACAGAAAGCAAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCAGACCTTTGA	3720
QY	3231	TTGATTTTAAAGACAATCTCACCTGTTTACCCAGGCTGGAGTGCAGTGGTGCATCA	3290	4311	GAGGCCACCTGGGAAGATTTGCTAGAATCTCAGGAGTTTCAAGACCCAGCCTGGGCAACACAGT	4370
Db	2641	TTGATTTTAAAGACAATCTCACCTGTTTACCCAGGCTGGAGTGCAGTGGTGCATCA	2700	3721	GAGGCCACCTGGGAAGATTTGCTAGAATCTCAGGAGTTTCAAGACCCAGCCTGGGCAACACAGT	3780
				4371	GAAACTCCATTTCTCCACAAAAATGGGAAAAAAGGAAAAATTCAGTGGTGTGCTGTGG	4430

QY 900 CCATCTTCTACTGATCCAAATCAGGAGCAAGGCCGTTGGGGTACCTGGTGGGGGTGATGC 959
 Db 638 CCATCTTCTACTGATCCAAATCAGGAGCAAGGCCGTTGGGGTACCTGGTGGGGGTGATGC 579
 QY 960 TGTTCAGGGAGGAGGCCCAAGGCAAGCTCAAAATTTGAATCTGAAGGCCCAATGCAC 1017
 Db 578 TGTTCAGGGAGGAGGCCCAAGGCAAGCTCAAAATTTGAATCTGAAGGCCCAATGCAC 521

RESULT 4
 AAC55315
 ID AAC55315 standard; DNA; 87 BP.
 XX
 AC AAC55315;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 DE Human activation-induced cytidine deaminase exon 1 SEQ ID NO:11.
 XX
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; anti-allergic;
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; rosen disease; digeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; Igg subclass selection disorder; ds.
 OS
 XX Homo sapiens.

PN WO200058480-A1.
 XX
 XX 05-OCT-2000.
 XX
 XX 28-MAR-2000; 2000WO-JP01918.
 XX
 PR 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.
 XX
 XX (NISR) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.
 XX
 PI Honjo T, Muramatsu M;
 XX
 DR WPI; 2000-611715/58.
 XX
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -
 XX
 PS Claim 18; Page 150; 174pp; Japanese.

XX The present invention describes an activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the
 CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (Iga) deficiency
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, rosen
 CC disease, digeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC Ige disorder, and Igg subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders. The
 CC present sequence represents the exon 1 genomic DNA sequence of human AID.

SQ Sequence 87 BP; 28 A; 17 C; 23 G; 19 T; 0 other;
 Query Match 1.6%; Score 87; DB 21; Length 87;
 Best Local Similarity 100.0%; Pred. No. 2.1e-21;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1032 AGAGAACCATCATTAATTTGAAGTGAAGTATTTTCTGGCCCTGAGACTTGCAGGAGGCAAGA 1091
 Db 1 AGAGAACCATCATTAATTTGAAGTGAAGTATTTTCTGGCCCTGAGACTTGCAGGAGGCAAGA 60
 QY 1092 AGACACTCTGGACACCACCTATGGACAG 1118
 Db 61 AGACACTCTGGACACCACCTATGGACAG 87

RESULT 5
 AAC55312
 ID AAC55312 standard; cDNA; 2818 BP.
 XX
 AC AAC55312;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 DE Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.
 XX
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;
 KW immune related disease; allergy; allergic disease; anti-allergic;
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
 KW gene therapy; B cell associated immune system disorder; food allergy;
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
 KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
 KW drug allergy; allergic rhinitis; rosen disease; digeorge disease; AIDS;
 KW ataxia telangiectasia; common variable immunodeficiency disorder;
 KW major histocompatibility class II deficiency disease;
 KW auto immunodeficiency syndrome; Igg subclass selection disorder; ss.

XX Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 80..676
 FT /tag= a
 FT /product= "activation-induced cytidine deaminase"

XX WO200058480-A1.
 XX
 XX 05-OCT-2000.
 XX
 XX 28-MAR-2000; 2000WO-JP01918.
 XX
 PR 29-MAR-1999; 99JP-0087192.
 PR 24-JUN-1999; 99JP-0178999.
 PR 27-DEC-1999; 99JP-0371382.

XX (NISR) JAPAN TOBACCO INC.
 PA (HONJ/) HONJO T.

XX Honjo T, Muramatsu M;

XX WPI; 2000-611715/58.

XX P-PSDB; AAB24198.

XX Nucleic acid encoding activation induced cytidine deaminase, useful as
 PT a target for drug development for immune-related diseases including
 PT allergies -

XX Claim 3; Page 135-139; 174pp; Japanese.

XX The present sequence encodes human activation-induced cytidine deaminase
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
 CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and
 CC dermatological activities, and can be used in gene therapy. AID
 CC polynucleotides are useful in methods for identifying drugs for the

CC treatment of B cell associated immune system disorders, immunodeficiency
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
CC disease, DiGeorge disease, ataxia telangiectasia, common variable
CC immunodeficiency disorder, MHC (major histocompatibility class) class
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
CC encoding AID may be used for gene therapy and the antibodies to the AID
CC protein may be used for diagnosis and treatment of these disorders.
XX

SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;

Query Match 1.6%; Score 87; DB 21; Length 2818;

Best Local Similarity 100.0%; Pred. No. 1.1e-21;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 AGAGAACCATCAATTAATGAGTGAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGA 1091

Db 1 AGAGAACCATCAATTAATGAGTGAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGA 60

QY 1092 AGACACTCTGGACACCATATGCACAG 1118

Db 61 AGACACTCTGGACACCATATGCACAG 87

RESULT 6

ABK52229/C

ID ABK52229 standard; cDNA; 12718 BP.

XX AC ABK52229;

XX 13-AUG-2002 (first entry)

DT cDNA encoding human CYP27A1 protein.

DE Human; Cytochrome P450; Subfamily XXVIIA; single nucleotide polymorphism;

KW Steroid 27-Hydroxylase; Cerebrotendinous Xanthomatosis Polypeptide 1;

KW CYP27A1; SNP; drug screening; cerebrotendinous xanthomatosis;

KW chromosome 2q33-qter; gene; ss.

XX Homo sapiens.

OS

XX Key

FH Location/Qualifiers

FT variation

FT replace (564,C)

FT /tag= a

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 1 (PS1)"

FT 1001..1255

FT /tag= b

FT /number= 1

FT 1001..0719

FT /tag= c

FT /product= "Human CYP27A1 protein"

FT 1256..3409

FT /tag= d

FT /number= 1

FT replace (1342,A)

FT /tag= e

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 2 (PS2)"

FT 3410..3600

FT /tag= f

FT /number= 2

FT replace (3564,A)

FT /tag= g

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 3 (PS3)"

FT replace (3592,A)

FT /tag= h

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 4 (PS4)"

FT 3601..5910

FT intron

FT variation

FT replace (3611,A)

FT /tag= j

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 5 (PS5)"

FT replace (5770,G)

FT /tag= k

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 6 (PS6)"

FT 5911..6110

FT /tag= l

FT /number= 3

FT replace (5970,T)

FT /tag= m

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 7 (PS7)"

FT replace (5988,T)

FT /tag= n

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 8 (PS8)"

FT 6111..6240

FT /tag= o

FT /number= 3

FT 6241..6438

FT /tag= p

FT /number= 4

FT 6439..6612

FT /tag= q

FT /number= 4

FT 6613..6785

FT /tag= r

FT /number= 5

FT replace (6689,A)

FT /tag= s

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 9 (PS9)"

FT 6786..7709

FT /tag= t

FT /number= 5

FT replace (6843,T)

FT /tag= u

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 10 (PS10)"

FT 7710..7876

FT /tag= v

FT /number= 6

FT replace (7843,T)

FT /tag= w

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 11 (PS11)"

FT 7877..8068

FT /tag= x

FT /number= 6

FT replace (7931,C)

FT /tag= y

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 12 (PS12)"

FT replace (8003,C)

FT /tag= z

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 13 (PS13)"

FT 8069..8147

FT /tag= aa

FT /number= 7

FT 8148..8233

FT /tag= ab

FT /number= 7

FT 8234..8446

FT /tag= ac

FT /number= 8

FT replace (8441,T)

FT /tag= ad

XX 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 9880.
DE Human reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
KW Homo sapiens.
XX WO200155320-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.


```
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251886.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 9880; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
XX Sequence 84 BP; 28 A; 20 C; 19 G; 17 T; 0 other:
SQ
Query Match 1.2%; Score 67; DB 22; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.8e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTGTGATTTTAGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 5505
Db 73 TTGTGATTTTAGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 14
QY 5506 ACCTCAG 5512
Db 13 ACCTCAG 7
RESULT 9
ID ABA18401 standard; DNA; 95 BP.
XX
AC ABA18401;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polynucleotide SEQ ID NO 10732.
XX
XX Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antiskickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebrotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
```

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX
XX Disclosure: SEQ ID NO 10732; 1701pp + Sequence Listing; English.
PS
XX
XX The invention relates to novel genes (AB11004-ABA21534) and proteins
CC (AB114678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful

CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 95 BP; 19 A; 26 C; 22 G; 28 T; 0 other;
SQ
Query Match 1.2%; Score 67; DB 22; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 5505
Db 18 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 77
Qy 5506 ACCTCAG 5512
Db 78 ACCTCAG 84
RESULT 10
AB14482/c
ID AB14482 standard; DNA; 121 BP.
XX
AC AB14482;
XX
XX 23-JAN-2002 (first entry)
XX Human nervous system related polynucleotide SEQ ID NO 6813.
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.

PR	14-AUG-2000;	2000US-02245518;
PR	14-AUG-2000;	2000US-02245519;
PR	14-AUG-2000;	2000US-02252113;
PR	14-AUG-2000;	2000US-02252114;
PR	14-AUG-2000;	2000US-02252166;
PR	14-AUG-2000;	2000US-02252667;
PR	14-AUG-2000;	2000US-02252668;
PR	14-AUG-2000;	2000US-02252700;
PR	14-AUG-2000;	2000US-02252709;
PR	14-AUG-2000;	2000US-02252758;
PR	14-AUG-2000;	2000US-02252759;
PR	18-AUG-2000;	2000US-02262679;
PR	22-AUG-2000;	2000US-02266681;
PR	22-AUG-2000;	2000US-02266688;
PR	23-AUG-2000;	2000US-02271182;
PR	23-AUG-2000;	2000US-02270009;
PR	30-AUG-2000;	2000US-02289294;
PR	01-SEP-2000;	2000US-02292887;
PR	01-SEP-2000;	2000US-02293433;
PR	01-SEP-2000;	2000US-02293444;
PR	01-SEP-2000;	2000US-02293445;
PR	05-SEP-2000;	2000US-02295009;
PR	05-SEP-2000;	2000US-02295113;
PR	06-SEP-2000;	2000US-02304337;
PR	06-SEP-2000;	2000US-02304338;
PR	08-SEP-2000;	2000US-02320800;
PR	08-SEP-2000;	2000US-02320881;
PR	12-SEP-2000;	2000US-02319688;
PR	14-SEP-2000;	2000US-02323397;
PR	14-SEP-2000;	2000US-02323398;
PR	14-SEP-2000;	2000US-02323399;
PR	14-SEP-2000;	2000US-02324200;
PR	14-SEP-2000;	2000US-02324201;
PR	14-SEP-2000;	2000US-02330663;
PR	14-SEP-2000;	2000US-02330864;
PR	21-SEP-2000;	2000US-02330655;
PR	21-SEP-2000;	2000US-02342223;
PR	21-SEP-2000;	2000US-02342274;
PR	25-SEP-2000;	2000US-02349977;
PR	25-SEP-2000;	2000US-02349998;
PR	25-SEP-2000;	2000US-02354884;
PR	27-SEP-2000;	2000US-02358834;
PR	27-SEP-2000;	2000US-02358836;
PR	29-SEP-2000;	2000US-02363237;
PR	29-SEP-2000;	2000US-02363667;
PR	29-SEP-2000;	2000US-02363668;
PR	29-SEP-2000;	2000US-02363669;
PR	29-SEP-2000;	2000US-02363700;
PR	29-SEP-2000;	2000US-02399935;
PR	13-OCT-2000;	2000US-02399937;
PR	02-OCT-2000;	2000US-02368002;
PR	02-OCT-2000;	2000US-02370037;
PR	02-OCT-2000;	2000US-02370038;
PR	02-OCT-2000;	2000US-02370039;
PR	02-OCT-2000;	2000US-02370040;
PR	13-OCT-2000;	2000US-02399935;
PR	13-OCT-2000;	2000US-02399937;
PR	20-OCT-2000;	2000US-02409600;
PR	20-OCT-2000;	2000US-02417885;
PR	20-OCT-2000;	2000US-02417886;
PR	20-OCT-2000;	2000US-02417887;
PR	20-OCT-2000;	2000US-02418008;
PR	20-OCT-2000;	2000US-02418009;
PR	20-OCT-2000;	2000US-02418026;
PR	20-OCT-2000;	2000US-02422221;
PR	01-NOV-2000;	2000US-02446177;
PR	08-NOV-2000;	2000US-02464774;
PR	08-NOV-2000;	2000US-02464775;
PR	08-NOV-2000;	2000US-02464776;

PR	08-NOV-2000;	2000US-0246477
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523
PR	08-NOV-2000;	2000US-0246524
PR	08-NOV-2000;	2000US-0246535
PR	08-NOV-2000;	2000US-0246536
PR	08-NOV-2000;	2000US-0246527
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611
PR	08-NOV-2000;	2000US-0246613
PR	17-NOV-2000;	2000US-0249207
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250391.
PR	01-DEC-2000;	2000US-0251160.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251959.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA. Barash SC. Ruben SM;

WPI: 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Disclosure: SEO ID NO 6813: 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABAl1004-ABAl21534) and proteins (ABAl4678-ABAl8001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast; and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 121 BP; 34 A; 33 C; 31 G; 23 T; 0 other;

Qy Query Match 1.2%; Score 67; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
5446 TTTCTATTATTAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 5505
Dc 79 TTTCTATTATTAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 20

Qy 5506 ACCTCAG 5512

Dc 19 ACCTCAG 13

RESULT 11

AAK80307/c

ID AAK80307 standard; DNA; 132 BP.

XX AAK80307;

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35119.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX WO200157182-A2.

PD 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0245474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

XX Disclosure; SEQ ID NO 35119; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.

XX Sequence 132 BP; 36 A; 36 C; 37 G; 23 T; 0 other;

Query Match 1.2%; Score 67; DB 22; Length 132;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTGTATTTTATGATAGATGGGGTTTACCATGTTGGCAGCGTGGTCTCAAACTCCG 5505
Db 129 TTTGTATTTTATGATAGATGGGGTTTACCATGTTGGCAGCGTGGTCTCAAACTCCG 70
QY 5506 ACCTCAG 5512
|||||||

Db 69 ACCTCAG 63
RESULT 12
AAK66840/C
ID AAK66840 standard; DNA; 141 BP.
XX
AC AAK66840;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:21652.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 28-JUN-2000; 2000US-0214886.
PR 07-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232406.
PR 14-SEP-2000; 2000US-0232407.
PR 14-SEP-2000; 2000US-0232423.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 21652; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 141 BP; 40 A; 34 C; 38 G; 29 T; 0 other;

Query Match 1.2%; Score 67; DB 22; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTATAGATAGATGGGTTTACACATGTTGGCAGCGTGTCTCAACATCTCTG 5505
|||||
Db 106 TTTGTATTTTATAGATAGATGGGTTTACACATGTTGGCAGCGTGTCTCAACATCTCTG 47
|||||
QY 5506 ACCTCAG 5512
Db 46 ACCTCAG 40
|||||

RESULT 13
AAK68098
ID AAK68098 standard; DNA; 141 BP.
XX
AC AAK68098;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22910.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX

OS Homo sapiens.
PN WO200157182-A2.
XX
XX
PD
XX
PF
XX
XX
PR 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
(HUMA-) HUMAN GENOME SCI INC.

```
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
DR Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX Disclosure; SEQ ID NO 22910; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX Sequence 141 BP; 29 A; 38 C; 34 G; 40 T; 0 other;
SQ

Query Match 1.2%; Score 67; DB 22; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5446 TTTCTATTATTAGATGGGCTTCCACCATGTGGCCAGGCTGCTCAAACTCCTG 5505
Db ||||||| 36 TTTCTATTATTAGATGGGCTTCCACCATGTGGCCAGGCTGCTCAAACTCCTG 95

Qy 5506 ACCTCAG 5512
Db ||||| 96 ACCTCAG 102

RESULT 14
AAK68878/c
ID AAK68878 standard; DNA; 141 BP.
XX AC AAK68878;
XX DT 06-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23690.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01354.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
```


PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0249208.
PR 08-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis.
XX
XX
XX Disclosure; SEQ ID NO 23690; 3071pp + Sequence Listing; English.
PS
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For

CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 141 BP; 40 A; 34 C; 38 G; 29 T; 0 other;
Query Match 1.2%; Score 67; DB 22; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTTGTATTTTGTAGATGGGGTTTCCACCATGTTGGCCAGCTGCTCAAACTCCTG 5505
Db 106 TTTGTATTTTGTAGATGGGGTTTCCACCATGTTGGCCAGCTGCTCAAACTCCTG 47
QY 5506 ACCTCAG 5512
Db 46 ACCTCAG 40
RESULT 15
AAK70059/c
ID AAK70059 standard; DNA; 141 BP.
XX
AC AAK70059;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:24871.
DE
XX
KW Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251719.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-483426/52.
DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
PT Disclosure; SEQ ID NO 24871; 3071pp + Sequence Listing; English.
PS AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 141 BP; 40 A; 34 C; 38 G; 29 T; 0 other;

Query Match

1.2%; Score 67; DB 22; Length 141;

Best Local Similarity 100.0%; Pred. No. 2.5e-14;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5446 TTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCCTG 5505
 Db 106 TTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCCTG 47
 QY 5506 ACCTCAG 5512
 Db 46 ACCTCAG 40

Search completed: June 19, 2003, 11:52:01
 Job time : 1074 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 11:34:27 ; Search time 720 seconds
(without alignments)
11238.046 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagagaaacatcgttccca.....tcaaaactcctgacctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1042519 seqs, 733713590 residues

Word size : 0 2085038

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5514	100.0	5514	9	US-09-966-880A-9
2	4771	86.5	11204	9	US-09-966-880A-35
3	87	1.6	87	9	US-09-966-880A-11
4	87	1.6	2818	9	US-09-966-880A-7
5	68	1.2	8158	9	US-09-832-232-30
c 6	67	1.2	84	9	US-09-764-891-9880
7	67	1.2	1160	9	US-09-764-891-5684
c 8	67	1.2	5930	10	US-09-764-877-2384
9	67	1.2	11754	9	US-09-984-827-5
c 10	67	1.2	16086	10	US-09-764-877-2385
c 11	67	1.2	24028	9	US-09-764-891-9207
c 12	67	1.2	32170	9	US-10-074-095-1108
c 13	67	1.2	32170	10	US-09-764-860-1108
c 14	67	1.2	34668	9	US-09-900-449A-3
c 15	67	1.2	58837	10	US-09-982-091A-5
c 16	67	1.2	122186	9	US-09-563-728A-36
c 17	67	1.2	143306	10	US-09-729-920A-3
c 18	67	1.2	174566	9	US-10-020-141-1
19	66	1.2	251	9	US-10-092-154-1205

20	66	1.2	251	10	US-09-764-847-1205	Sequence 1205, Ap
21	66	1.2	285	9	US-10-091-504-1783	Sequence 1783, Ap
22	66	1.2	285	10	US-09-764-869-1783	Sequence 1783, Ap
23	66	1.2	638	9	US-10-091-572-628	Sequence 628, App
24	66	1.2	638	9	US-09-764-851-6781	Sequence 6781, Ap
25	66	1.2	6257	9	US-10-091-504-1782	Sequence 1782, Ap
26	66	1.2	6257	10	US-09-764-869-1782	Sequence 1782, Ap
27	66	1.2	6761	9	US-10-091-504-1724	Sequence 1724, Ap
28	66	1.2	6761	10	US-09-764-869-1724	Sequence 1724, Ap
c 29	66	1.2	7219	9	US-10-102-627-107	Sequence 107, App
30	66	1.2	8758	10	US-09-962-276-3	Sequence 3, Appli
31	66	1.2	12118	10	US-09-788-654A-3	Sequence 3, Appli
32	66	1.2	14093	9	US-10-092-154-1744	Sequence 1744, Ap
33	66	1.2	14093	10	US-09-764-847-1744	Sequence 1744, Ap
34	66	1.2	28588	9	US-10-073-961-399	Sequence 399, App
35	66	1.2	28588	10	US-09-764-887-399	Sequence 399, App
36	66	1.2	32146	9	US-10-074-095-797	Sequence 797, App
37	66	1.2	32146	10	US-09-764-860-797	Sequence 797, App
38	66	1.2	32248	9	US-10-074-095-802	Sequence 802, App
39	66	1.2	32248	10	US-09-764-860-802	Sequence 802, App
40	66	1.2	52216	10	US-09-747-810-1	Sequence 1, Appli
41	65	1.2	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c 42	64	1.2	426	9	US-09-918-995-4335	Sequence 4335, Ap
c 43	64	1.2	3569	9	US-10-091-572-737	Sequence 737, App
c 44	64	1.2	3569	9	US-09-764-891-9161	Sequence 9161, Ap
45	64	1.2	10527	9	US-10-116-420-100	Sequence 100, App

ALIGNMENTS

RESULT 1

US-09-966-880A-9
; Sequence 9, Application US/099666880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasaku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JF00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 5514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1)...(1031)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1032)...(1118)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1119)...(5514)
; US-09-966-880A-9

Query Match 100.0%; Score 5514; DB 9; Length 5514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACAGACGAATACATGGTCCAAAGCTAGGCTATTGATTTGAAATCATCAAGGTATAGATG 60
|||||

Db	1	ACAGACGAATACATGGTCCAAAGCTAGGGCTATTGATTTGAAATCATCAAGGTATAGATG	60
Qy	61	GTATCAAAAGGCTTGAGGCAGGAAGAGAGACAGAGACCCCTAGCTGCATGCTTAGCATTGCA	120
Db	61	GTATCAAAAGGCTTGAGGCAGGAAGAGAGACAGAGACCCCTAGCTGCATGCTTAGCATTGCA	120
Qy	121	TCCCTAGCACCTGGCATAGTTTCCCAATTAACAGTAGGCATGAAGTATCTACTCAGTGAATA	180
Db	121	TCCCTAGCACCTGGCATAGTTTCCCAATTAACAGTAGGCATGAAGTATCTACTCAGTGAATA	180
Qy	181	AATAGAAATGCATATGGGCTACAGTAGGAGAGAGAAATAAAATCTTTAATAGACCAGTTC	240
Db	181	AATAGAAATGCATATGGGCTACAGTAGGAGAGAGAAATAAAATCTTTAATAGACCAGTTC	240
Qy	241	TATGAGAGACAAAATTAAGTCTTTTATTTCGAAGATCTTAGCCCTGTTTCCCAATTCAG	300
Db	241	TATGAGAGACAAAATTAAGTCTTTTATTTCGAAGATCTTAGCCCTGTTTCCCAATTCAG	300
Qy	301	TGCAGCCAGTTAGACACATGATCTGTCTGGTGAACAAGCAATTTTGTATTTTGGGGGAC	360
Db	301	TGCAGCCAGTTAGACACATGATCTGTCTGGTGAACAAGCAATTTTGTATTTTGGGGGAC	360
Qy	361	TGCTGCTGCTTCTGACTCCAAATTAAGGATTTTTTTTTTCTAAAAAGATGGCTCATG	420
Db	361	TGCTGCTGCTTCTGACTCCAAATTAAGGATTTTTTTTTTCTAAAAAGATGGCTCATG	420
Qy	421	CAAAAATCACTCTTTGGTCTAAATATCTAGTCTTCAAGCAATCTTGTATGCAATCAGA	480
Db	421	CAAAAATCACTCTTTGGTCTAAATATCTAGTCTTCAAGCAATCTTGTATGCAATCAGA	480
Qy	481	AAGAAAAATCCATGGTTTGGAGGCAAAAATTTTGTGTTCTTAAATCTATATAACTGA	540
Db	481	AAGAAAAATCCATGGTTTGGAGGCAAAAATTTTGTGTTCTTAAATCTATATAACTGA	540
Qy	541	GTTTCATTTGCTTAACTGCAAGACAGAGCTGTAGTGCCTGTCTGACTGAGGTTCCAGAG	600
Db	541	GTTTCATTTGCTTAACTGCAAGACAGAGCTGTAGTGCCTGTCTGACTGAGGTTCCAGAG	600
Qy	601	AGACTGTGGGAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACACAATAAACC	660
Db	601	AGACTGTGGGAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACACAATAAACC	660
Qy	661	GAAGCTATTTAAATGCTCTTTAAGGTATTTAGATAAATAATTAATCTATCTCAATTTGCTTT	720
Db	661	GAAGCTATTTAAATGCTCTTTAAGGTATTTAGATAAATAATTAATCTATCTCAATTTGCTTT	720
Qy	721	TATTTTGTCTATCATGATTAATAATGAAGTGTCTACTCTTACTGCTCTCTGATCTTTGC	780
Db	721	TATTTTGTCTATCATGATTAATAATGAAGTGTCTACTCTTACTGCTCTCTGATCTTTGC	780
Qy	781	TAGCTATGAGCATGGACTGGGCTTTTAGAGCAGACGCCCAAGGAACCTTAAACATTA	840
Db	781	TAGCTATGAGCATGGACTGGGCTTTTAGAGCAGACGCCCAAGGAACCTTAAACATTA	840
Qy	841	AGCAGAGTGCCTCAATGGTTTAACTGTGTGACTCTGCCTATGACAGCCCAACCCACC	900
Db	841	AGCAGAGTGCCTCAATGGTTTAACTGTGTGACTCTGCCTATGACAGCCCAACCCACC	900
Qy	901	CATCTTCACTGGATCCAAATCAGGACAGGCGGTTGGGGTACCTGGTGGGGTGATGCT	960
Db	901	CATCTTCACTGGATCCAAATCAGGACAGGCGGTTGGGGTACCTGGTGGGGTGATGCT	960
Qy	961	GTGAGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGGCAATGCACTGT	1020
Db	961	GTGAGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGGCAATGCACTGT	1020
Qy	1021	CAGACTGACACAGAACCATCATTAATTTGAAGTGAGATTTTCTGGCTGAGACTTGCA	1080
Db	1021	CAGACTGACACAGAACCATCATTAATTTGAAGTGAGATTTTCTGGCTGAGACTTGCA	1080
Qy	1081	GGGAGCAAGAAGACACTCTGGACACCCTATGGACAGGTAAAGAGGCGAGTCTTCTCGTG	1140
Db	1081	GGGAGCAAGAAGACACTCTGGACACCCTATGGACAGGTAAAGAGGCGAGTCTTCTCGTG	1140

Qy	1141	GGTGATTCGACATGGGCTTCCTCTCAGAGCAAAATCTGAGTAATGAGACTGGTAGCTATGCC	1200
Db	1141	GGTGATTCGACATGGGCTTCCTCTCAGAGCAAAATCTGAGTAATGAGACTGGTAGCTATGCC	1200
Qy	1201	TTTCTCTCATGTAACACTGCTGACTGATAAGATCAGCTTGATCAATATGATATATATTTT	1260
Db	1201	TTTCTCTCATGTAACACTGCTGACTGATAAGATCAGCTTGATCAATATGATATATATATTTT	1260
Qy	1261	TTGATCTGCTCCTCTTTCTTCTATTCAGATCTTATACGCTGTGAGCCCAATTTCTTCTGT	1320
Db	1261	TTGATCTGCTCCTCTTTCTTCTATTCAGATCTTATACGCTGTGAGCCCAATTTCTTCTGT	1320
Qy	1321	TTTCTCTCTCTTGTGATTTCCCTCTTTTTCATGTGGCAAAAGTAGTGGCGTACAAATGT	1380
Db	1321	TTTCTCTCTCTTGTGATTTCCCTCTTTTTCATGTGGCAAAAGTAGTGGCGTACAAATGT	1380
Qy	1381	ACTGATTCGCTGAGATTTGTACCATGTTTGAACCTAATTTATGTTAATAATATTAACA	1440
Db	1381	ACTGATTCGCTGAGATTTGTACCATGTTTGAACCTAATTTATGTTAATAATATTAACA	1440
Qy	1441	TAGCAAACTCTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTACTAAAAACGG	1500
Db	1441	TAGCAAACTCTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTACTAAAAACGG	1500
Qy	1501	TAGTGTAAATTTTCGTAATTTTGTAAATATTTCAACAGTAAACCAACTTGAAGACACA	1560
Db	1501	TAGTGTAAATTTTCGTAATTTTGTAAATATTTCAACAGTAAACCAACTTGAAGACACA	1560
Qy	1561	CTTTCCTAGGAGGCGTTACTGAAATTAATTTAGCTATAGTAAGAAAATTTGTAATTTAG	1620
Db	1561	CTTTCCTAGGAGGCGTTACTGAAATTAATTTAGCTATAGTAAGAAAATTTGTAATTTAG	1620
Qy	1621	AAATGCCAAGCATTTTAAATTTAATGCTTGAAGTCACTATGATGTGTCCATTTAAGG	1680
Db	1621	AAATGCCAAGCATTTTAAATTTAATGCTTGAAGTCACTATGATGTGTCCATTTAAGG	1680
Qy	1681	AGCAAAATTCATCAAGCAAGTTTAAATTTAAGTAAAGGCCCAATTTAGGCAAGTTAATG	1740
Db	1681	AGCAAAATTCATCAAGCAAGTTTAAATTTAAGTAAAGGCCCAATTTAGGCAAGTTAATG	1740
Qy	1741	GCATTTTACTATTAACCTAATCTTCCATTTGTTTCAAGCTAGCTTAACTTACCTCTTAG	1800
Db	1741	GCATTTTACTATTAACCTAATCTTCCATTTGTTTCAAGCTAGCTTAACTTACCTCTTAG	1800
Qy	1801	GTGTGAATTTGGTTAAGGCTCCTCATATATGCTTTTATGTGCAGTTTTGATAGTATTGT	1860
Db	1801	GTGTGAATTTGGTTAAGGCTCCTCATATATGCTTTTATGTGCAGTTTTGATAGTATTGT	1860
Qy	1861	CATAGAACTTATTCCTACATTTTATGATTTACTATGGATGTATGAGAAATACACCTA	1920
Db	1861	CATAGAACTTATTCCTACATTTTATGATTTACTATGGATGTATGAGAAATACACCTA	1920
Qy	1921	ATCCTTATCTTTACCTCAATTTAATCTTCTTAAAGAACTTACATTTACAGAAATAAGA	1980
Db	1921	ATCCTTATCTTTACCTCAATTTAATCTTCTTAAAGAACTTACATTTACAGAAATAAGA	1980
Qy	1981	TTTTTTAAAAATATTTTTTTTGTAGACAGAGGCTTTAGCCAGCCGAGGCTGGTCTCT	2040
Db	1981	TTTTTTAAAAATATTTTTTTTGTAGACAGAGGCTTTAGCCAGCCGAGGCTGGTCTCT	2040
Qy	2041	AAGTCTGGCCCAAGGATCCTCCTGGCTGGGCTCCTAAAGTCTGCGAATTTATAGACAT	2100
Db	2041	AAGTCTGGCCCAAGGATCCTCCTGGCTGGGCTCCTAAAGTCTGCGAATTTATAGACAT	2100
Qy	2101	GAGCCATCACATCCAATATACAGAAATAAGATTTTAAATGGAGGATTTAATGTTCTTCAG	2160
Db	2101	GAGCCATCACATCCAATATACAGAAATAAGATTTTAAATGGAGGATTTAATGTTCTTCAG	2160
Qy	2161	AAAAATTTCTTTGAGGTCAGACAAATGTCTCTCAGTTTACACTGAGATTTTGA	2220
Db	2161	AAAAATTTCTTTGAGGTCAGACAAATGTCTCTCAGTTTACACTGAGATTTTGA	2220

Db 4381 TCTCCACAAAATGGAAAAAAGAAAGCAAAATCAGTGGTTCTCTGTGGGAGGGGAAG 4440

QY 4441 GACTCAAGAGAGGAGAGCTCTGTGGGGTGAGGGTGGTGAATTCAGGTTCTGTATCTCT 4500

Db 4441 GACTCAAGAGAGGAGAGCTCTGTGGGGTGAGGGTGGTGAATTCAGGTTCTGTATCTCT 4500

QY 4501 GACTGTGGTAGCAGTTTGGGGTGTACATCCAAATAATTCGTAGAATATGCAATCTTA 4560

Db 4501 GACTGTGGTAGCAGTTTGGGGTGTACATCCAAATAATTCGTAGAATATGCAATCTTA 4560

QY 4561 AATGGGTGCAGTTTACTGTATCAATTAATCACTCAATGTAAGAAAAATAATGCTAAG 4620

Db 4561 AATGGGTGCAGTTTACTGTATCAATTAATCACTCAATGTAAGAAAAATAATGCTAAG 4620

QY 4621 AAAAGTTTCAATTCCTTCCAGCAACGTTATTCAAATTCCTGAGCCCTTTACTTCGCA 4680

Db 4621 AAAAGTTTCAATTCCTTCCAGCAACGTTATTCAAATTCCTGAGCCCTTTACTTCGCA 4680

QY 4681 AATTCCTGCACTTCTGCCCCGTACCATPAGGTGACAGCACTAGCTCCACAAATGGATA 4740

Db 4681 AATTCCTGCACTTCTGCCCCGTACCATPAGGTGACAGCACTAGCTCCACAAATGGATA 4740

QY 4741 AATGCATTTCTGAAAAGACTAGGCAAAATCCAGGCATCACTGTGCTTCATATCAA 4800

Db 4741 AATGCATTTCTGAAAAGACTAGGCAAAATCCAGGCATCACTGTGCTTCATATCAA 4800

QY 4801 CCACGCTGTACAGCTTGTGCTGTGCTGCTGCTGCTGCAATGGGACTCTTTGATTTTAA 4860

Db 4801 CCACGCTGTACAGCTTGTGCTGTGCTGCTGCTGCTGCAATGGGACTCTTTGATTTTAA 4860

QY 4861 GGAACCTGGGTTACAGAGTATTTCCACAAATGCTATTCAAATAGTGCCTTATGATAG 4920

Db 4861 GGAACCTGGGTTACAGAGTATTTCCACAAATGCTATTCAAATAGTGCCTTATGATAG 4920

QY 4921 CAAGACACTGCTTAGGAGCCAGAAACAAAGGAGGAGAAATCAGTCATTTATGCGCA 4980

Db 4921 CAAGACACTGCTTAGGAGCCAGAAACAAAGGAGGAGAAATCAGTCATTTATGCGCA 4980

QY 4981 ACAACATGCAAGATATTTAGATCAITTTTGACTAGTTTAAAAAGCAGCAGAGTACAAAAT 5040

Db 4981 ACAACATGCAAGATATTTAGATCAITTTTGACTAGTTTAAAAAGCAGCAGAGTACAAAAT 5040

QY 5041 CACACATGCAATCAGTATTAATCCAAATCATGTAATATGTCCTGTAGAAAGACTAGAGG 5100

Db 5041 CACACATGCAATCAGTATTAATCCAAATCATGTAATATGTCCTGTAGAAAGACTAGAGG 5100

QY 5101 AATAAACACAAGAATCTTAACAGTCAATTCATTTAGACACTAAGTCTAATTTATTTATTT 5160

Db 5101 AATAAACACAAGAATCTTAACAGTCAATTCATTTAGACACTAAGTCTAATTTATTTATTT 5160

QY 5161 AGACACTATGATATTTGAGATTTAAAAATCTTTTAAATTTTAAAAATTTAGAGCTCTCT 5220

Db 5161 AGACACTATGATATTTGAGATTTAAAAATCTTTTAAATTTTAAAAATTTAGAGCTCTCT 5220

QY 5221 ATTTTCCATAGTATTCAGTTTGCAATGATCAAGTATTAATCTCTCTTTTTTTTTTTTT 5280

Db 5221 ATTTTCCATAGTATTCAGTTTGCAATGATCAAGTATTAATCTCTCTTTTTTTTTTTTT 5280

QY 5281 TTTTTTTTTTTTGGAGATGGAGTTTGGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5340

Db 5281 TTTTTTTTTTTTGGAGATGGAGTTTGGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5340

QY 5341 AYCATACTGCTGCAACCTCCACCTCTCTGGGTTCAAGAAAGCTGTCGCCCTCAGCCTCC 5400

Db 5341 AYCATACTGCTGCAACCTCCACCTCTCTGGGTTCAAGAAAGCTGTCGCCCTCAGCCTCC 5400

QY 5401 CGGGTAGATGGGATTTACAGCGCCGCCACACCTCGGCTAATGTTTGTATTTTATTTAGTA 5460

Db 5401 CGGGTAGATGGGATTTACAGCGCCGCCACACCTCGGCTAATGTTTGTATTTTATTTAGTA 5460

QY 5461 GAGATGGGTTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAG 5514

Db 5461 GAGATGGGTTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAG 5514

RESULT 2

US-09-966-880A-35

; Sequence 35, Application US/09966880A

; Patent No. US2002016473A1

; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku

; APPLICANT: Muramatsu, Masamichi

; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE

; FILE REFERENCE: 06501-088001

; CURRENT APPLICATION NUMBER: US/09/966, 880A

; PRIOR FILING DATE: 2001-09-28

; PRIOR APPLICATION NUMBER: PCT/JP00/01918

; PRIOR FILING DATE: 2000-03-28

; PRIOR FILING DATE: 1999-12-27

; PRIOR FILING DATE: 1999-12-27

; PRIOR APPLICATION NUMBER: JP 11-371382

; PRIOR FILING DATE: 1999-06-24

; PRIOR APPLICATION NUMBER: JP 11-87192

; PRIOR FILING DATE: 1999-03-29

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 35

; LENGTH: 11204

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-966-880A-35

Query Match 86.5%; Score 4771; DB 9; Length 11204;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 4921; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 591 AGTTTCAGAGAGACTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTCAACAC 650

Db 1 AGTTTCAGAGAGACTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTCAACAC 60

QY 651 AATAACCCCAAGAGCTATTTAAATGCTCTTTAAGGTATTTACATAAATATTACTATTCTC 710

Db 61 AATAACCCCAAGAGCTATTTAAATGCTCTTTAAGGTATTTACATAAATATTACTATTCTC 120

QY 711 ATTGCTCTTTATTGTTGTTATCATGATTATTAATGAAGTGCTACTGTTACTGCTCTCC 770

Db 121 ATTGCTCTTTATTGTTGTTATCATGATTATTAATGAAGTGCTACTGTTACTGCTCTCC 180

QY 771 TGATCTTTGCTAGCTATGGAGCATGGAGCTGGGCTTTTACAGCAGCAGCCCCAAAGGAACC 830

Db 181 TGATCTTTGCTAGCTATGGAGCATGGAGCTGGGCTTTTACAGCAGCAGCCCCAAAGGAACC 240

QY 831 TAAACATTTAAGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCTATGACAGC 890

Db 241 TAAACATTTAAGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCTATGACAGC 300

QY 891 CCACCCACCCTCTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGTACTCTGGTGG 950

Db 301 CCACCCACCCTCTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGTACTCTGGTGG 360

QY 951 GGGTGATGCTGTGTCAGGGGAGGAGCCCAAAAGGCAAGCTCAAAATTTGAATGTGAAGGCC 1010

Db 361 GGGTGATGCTGTGTCAGGGGAGGAGCCCAAAAGGCAAGCTCAAAATTTGAATGTGAAGGCC 420

QY 1011 AATGCACTGTGACACTGTAGACAGAGAACCATCATTAATTTGAAGTGAGATTTTCTGGCCT 1070

Db 421 AATGCACTGTGACACTGTAGACAGAGAACCATCATTAATTTGAAGTGAGATTTTCTGGCCT 480

QY 1071 GAGACTTGCAGGGAGGCAAGACACTCTGGACACACTATGGACAGCTATGGACAGTAAAGAGCAG 1130

Db 481 GAGACTTGCAGGGAGGCAAGACACTCTGGACACACTATGGACAGCTATGGACAGTAAAGAGCAG 540

QY 1131 TCTTCTCGGGTGATTTGCACTGGCCTTCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 1190

Db 541 TCTTCTCGGGTGATTTGCACTGGCCTTCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 600


```
RESULT 3
US-09-966-880A-11
; Sequence 11, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966.880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 87
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-11

Query Match      1.6%; Score 87; DB 9; Length 87;
Best Local Similarity 100.0%; Pred. No. 8.7e-31;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 1091
Db 1 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 60

QY 1092 AGACACTCTGGACACCACTATGGACAG 1118
Db 61 AGACACTCTGGACACCACTATGGACAG 87

RESULT 4
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966.880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(79)
; FEATURE:
```

```
; NAME/KEY: 3'UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

Query Match      1.6%; Score 87; DB 9; Length 2818;
Best Local Similarity 100.0%; Pred. No. 1.1e-30;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 1091
Db 1 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 60

QY 1092 AGACACTCTGGACACCACTATGGACAG 1118
Db 61 AGACACTCTGGACACCACTATGGACAG 87

RESULT 5
US-09-832-292-30
; Sequence 30, Application US/09832292
; Patent No. US20020177205A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey
; TITLE OF INVENTION: MAMMALIAN ALPHA-KINASE PROTEINS, NUCLEIC ACIDS AND
; FILE REFERENCE: 601-1-098CIP
; CURRENT APPLICATION NUMBER: US/09/832.292
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 09/632,131
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 8158
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-832-292-30

Query Match      1.2%; Score 68; DB 9; Length 8158;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCTG 5505
Db 6996 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCTG 7055

QY 5506 ACCTCAGA 5513
Db 7056 ACCTCAGA 7063

RESULT 6
US-09-764-891-9880/c
; Sequence 9880, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764.891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9880
; LENGTH: 84
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9880

Query Match      1.2%; Score 67; DB 9; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.7e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
 |||||||
 Db 73 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 14
 |||||||
 QY 5506 ACCTCAG 5512
 |||||||
 Db 13 ACCTCAG 7
 |||||||
 RESULT 7
 US-09-764-891-5684
 ; Sequence 5684, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5684
 ; LENGTH: 1160
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-891-5684

Query Match 1.2%; Score 67; DB 9; Length 1160;
 Best Local Similarity 100.0%; Pred. No. 3.1e-21;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
 |||||||
 Db 170 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 229
 |||||||
 QY 5506 ACCTCAG 5512
 |||||||
 Db 230 ACCTCAG 236
 |||||||

RESULT 8
 US-09-764-877-2384/c
 ; Sequence 2384, Application US/09764877
 ; Patent No. US20020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC005
 ; CURRENT APPLICATION NUMBER: US/09/764,877
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 4031
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2384
 ; LENGTH: 5930
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-877-2384

Query Match 1.2%; Score 67; DB 10; Length 5930;
 Best Local Similarity 100.0%; Pred. No. 3.4e-21;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
 |||||||
 Db 1254 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 1195
 |||||||
 QY 5506 ACCTCAG 5512
 |||||||
 Db 1194 ACCTCAG 1188
 |||||||

RESULT 9
 US-09-984-827-5
 ; Sequence 5, Application US/0984827
 ; Publication No. US20030056234A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DENEFELE, PATRICE
 ; APPLICANT: ROSIER-MONTUS, MARIE-FRANCOISE
 ; APPLICANT: ARNOULD-REGUIGNE, ISABELLE
 ; APPLICANT: DUVERGER, NICOLAS
 ; APPLICANT: CAMBIEN, FRANCOIS
 ; TITLE OF INVENTION: POLYMORPHIC SEQUENCES OF THE HUMAN ABCA1 GENE, THEIR USES, AND
 ; FILE REFERENCE: 03806.0522-00000
 ; CURRENT APPLICATION NUMBER: US/09/984,827
 ; CURRENT FILING DATE: 2002-04-01
 ; PRIOR APPLICATION NUMBER: 60/254,108
 ; PRIOR FILING DATE: 2000-12-11
 ; PRIOR APPLICATION NUMBER: FR 00/14037
 ; PRIOR FILING DATE: 2000-10-31
 ; NUMBER OF SEQ ID NOS: 161
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 11754
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-984-827-5

Query Match 1.2%; Score 67; DB 9; Length 11754;
 Best Local Similarity 100.0%; Pred. No. 3.6e-21;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
 |||||||
 Db 2032 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 2091
 |||||||
 QY 5506 ACCTCAG 5512
 |||||||
 Db 2092 ACCTCAG 2098
 |||||||

RESULT 10
 US-09-764-877-2385/c
 ; Sequence 2385, Application US/09764877
 ; Patent No. US20020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC005
 ; CURRENT APPLICATION NUMBER: US/09/764,877
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 4031
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2385
 ; LENGTH: 16086
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-877-2385

Query Match 1.2%; Score 67; DB 10; Length 16086;
 Best Local Similarity 100.0%; Pred. No. 3.6e-21;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
 |||||||
 Db 11409 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 11350
 |||||||
 QY 5506 ACCTCAG 5512
 |||||||
 Db 11349 ACCTCAG 11343
 |||||||

```
RESULT 11
US-09-764-891-8207/c
; Sequence 8207, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 8207
; LENGTH: 24028
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8207

Query Match          1.2%  Score 67;  DB 9;  Length 24028;
Best Local Similarity 100.0%;  Pred. No. 3.7e-21;
Matches 67;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  5446  TTTGTTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 5505
      |||||||
Db  19797  TTTGTTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 19738

Qy  5506  ACCTCAG 5512
      |||||||
Db  19737  ACCTCAG 19731

RESULT 12
US-10-074-095-1108/c
; Sequence 1108, Application US/10074095
; Publication No. US20030077704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 09/764,860
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
```

;
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399

;
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match 1.2%; Score 67; DB 9; Length 32170;
Best Local Similarity 100.0%; Pred. No. 3.8e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 5505
|||||
Db 28881 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 28822
|||||

QY 5506 ACCTCAG 5512
|||||
Db 28821 ACCTCAG 28815
|||||

RESULT 13
US-09-764-860-1108/c
; Sequence 1108, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1108
; LENGTH: 32170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-1108

Query Match 1.2%; Score 67; DB 10; Length 32170;
Best Local Similarity 100.0%; Pred. No. 3.8e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 5505
|||||
Db 28881 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 28822
|||||

QY 5506 ACCTCAG 5512
|||||
Db 28821 ACCTCAG 28815
|||||

RESULT 14
US-09-900-449A-3/c
; Sequence 3, Application US/09900449A
; Publication No. US20030040616A1
; GENERAL INFORMATION:
; APPLICANT: ZHONG, Jenny et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001271
; CURRENT APPLICATION NUMBER: US/09/900,449A
; CURRENT FILING DATE: 2001-07-09

; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 34668
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-900-449A-3

Query Match 1.2%; Score 67; DB 9; Length 34668;
Best Local Similarity 100.0%; Pred. No. 3.8e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTAATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 5505
|||||
DB 23213 TTTGTAATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 23154

QY 5506 ACCTCAG 5512
|||||
DB 23153 ACCTCAG 23147

RESULT 15

US-09-982-091A-5
; Sequence 5, Application US/09982091A
; Patent No. US20020151030A1
; GENERAL INFORMATION:
; APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
; APPLICANT: KUMAGAI, Akiko
; APPLICANT: DUNPHY, William
; TITLE OF INVENTION: CLASPIN PROTEINS AND METHODS OF USE THEREOF
; FILE REFERENCE: CITI320-1
; CURRENT APPLICATION NUMBER: US/09/982.091A
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 60/241,246
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 58837
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-982-091A-5

Query Match 1.2%; Score 67; DB 10; Length 58837;
Best Local Similarity 100.0%; Pred. No. 3.9e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTAATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 5505
|||||
DB 55194 TTTGTAATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 55253

QY 5506 ACCTCAG 5512
|||||
DB 55254 ACCTCAG 55260

Search completed: June 19, 2003, 17:54:07
Job time : 722 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:30:49 ; Search time 221.5 Seconds
(without alignments)
6634.257 Million cell updates/sec

Title: US-09-966-880A-9

Perfect score: 1736

Sequence: 1 acagacgaatacatagtggtcca.....tcaaaactcctgacctcagag 5514

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-Q=/cyn2_1/USPrto_spool/US09966880/runat_14062003_175620_11018/app_query.fasta_1.5703
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=olin2p.rag -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -NATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 -CGCN_1_1_379 @runat_14062003_175620_11018 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEPOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A_Geneseq_101002.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	39	2.2	251	22	ABG17006
c 2	22	1.3	75	22	AAO02240
c 3	22	1.3	103	21	AAB43997
4	22	1.3	110	21	AAB43292
5	22	1.3	116	22	AAB93273
6	22	1.3	137	22	ABG19377
7	22	1.3	407	22	AAU30450
8	21	1.2	80	22	AAU86859
9	21	1.2	118	22	AAU93908
10	20	1.2	36	22	AAO12441
11	20	1.2	42	22	ABG11938
12	20	1.2	89	22	ABG09800
13	20	1.2	125	22	AAU41856
c 14	19	1.1	41	22	AAO05250
15	19	1.1	55	22	AAO09035
16	19	1.1	93	22	AAU32720
c 17	19	1.1	110	23	ABP41552
18	19	1.1	115	22	AAU6287
19	19	1.1	115	22	AAU42316
c 20	19	1.1	115	23	ABG68710
c 21	19	1.1	136	23	AAE14727
c 22	19	1.1	141	20	AAU19767
23	19	1.1	188	22	ABG11932
c 24	18	1.0	44	22	AAO05621
25	18	1.0	60	22	AAU86602
c 26	18	1.0	63	22	AAO02094
c 27	18	1.0	81	22	AAU62666
28	18	1.0	81	22	AAU42307
29	18	1.0	86	21	AAU42412
c 30	18	1.0	90	22	AAU30185
c 31	18	1.0	99	22	AAU78806
c 32	18	1.0	104	22	AAO01382
c 33	18	1.0	106	22	AAU08280
c 34	18	1.0	131	22	AAU29521
c 35	17	1.0	25	22	ABG08228
36	17	1.0	33	22	AAU08563
37	17	1.0	34	22	AAU31173
c 38	17	1.0	35	22	AAO11772
c 39	17	1.0	37	22	AAO05588
c 40	17	1.0	37	22	AAO13714
41	17	1.0	38	22	AAO13748
42	17	1.0	39	22	AAO03987
43	17	1.0	40	22	AAO07106
44	17	1.0	43	22	AAU62290
c 45	17	1.0	45	22	ABG05119

ALIGNMENTS

RESULT 1
ABG17006
ID ABG17006 standard; Protein; 251 AA.
XX
AC ABG17006;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16997.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
-PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS81193.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX

PS Claim 20; SEQ ID No 47365; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 251 AA;

SQ Alignment Scores:

Pred. No.: 3,1e-28 Length: 251
Score: 39,00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2,24% Indels: 0
DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x ABG17006 (1-251)

QY 1017 GTGCATTGGCCCTCACATTCAAAATTGAGCTTGCCCTTTGGGCTCTCCCTGCACAC 958
DB 169 ValHisTrpProPheThrPhenylsPheGluLeuAlaLeuLeuGlySerProAspSer 188
QY 957 ATCACCCACACAGGTACCCCAAGCGCTTGCTCTGATTGGATCCAGTGAAGATG 901
DB 189 IleThrProThrArgTyrProAsnGlyLeuAlaProAspLeuAspProVallylMet 207

RESULT 2

AAO02240
ID AAO02240 standard; Protein; 75 AA.

XX AAO02240;

XX 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 16132.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-514838/56.

DR N-PSDB; AAI82171.

XX Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

PS Claim 20; SEQ ID NO 16132; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 75 AA;

SQ Alignment Scores:

Pred. No.: 9,03e-12 Length: 75
Score: 22,00 Matches: 22
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1,27% Indels: 0
DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x AAO02240 (1-75)

QY 3623 TCAAGGGATCCACCCACCTCAGCTCCAGAGTGCTGGGATTACAGGCGTGAGCCACCAC 3564
DB 35 SerSerAspProProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 54

QY 3563 ACCTGG 3558

DB 55 ThrTrp 56

RESULT 3

AA043997
ID AAB43997 standard; Protein; 103 AA.

XX AAB43997;

XX 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO:1442.

XX Human; cancer associated gene; cancer antigen; detection; cancer;
KW

diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 antidiabetic; antiaesthatic; antirheumatic; antiarthritic; antiviral;
 antinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
 dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 vasotrophic; antipsoriatic; angiogenic; gene therapy; inflammation;
 immune disorder; haematopoietic cell disorder; autoimmune disorder;
 allergic reaction; graft versus host disease; organ rejection;
 haemostatic; thrombolytic; cardiovascular disorder; infection;
 neurological disease; drug screening.
 XX Homo sapiens.
 OS
 XX WO200055350-A1.
 PN
 XX 21-SEP-2000.
 PD
 XX 08-MAR-2000; 2000WO-US05882.
 PF
 XX 12-MAR-1999; 99US-0124270.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM;
 PI
 XX WPI; 2000-587533/55.
 DR
 XX N-PSDB; AAC78206.
 DR
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 useful for treating or diagnosing e.g. cancer -
 PT
 XX
 XX Claim 11; Page 2126; 2352pp; English.
 PS
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given
 in AAB43398 to AAB44239. The proteins can have activities based on the
 tissues and cells the genes are expressed in. Example of activities
 include: cytostatic; proliferative; vulnery; immunomodulator;
 antidiabetic; antiaesthatic; antirheumatic; antiarthritic;
 antinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 nootropic; vasotrophic; antipsoriatic and angiogenic. The
 polynucleotides and polypeptides can be used for preventing, treating or
 ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 the present invention may be used to treat immune disorders by activating
 or inhibiting the proliferation, differentiation or mobilisation of
 immune cells, to treat disorders of haematopoietic cells, autoimmune
 disorders, allergic reactions, graft versus host disease and organ
 rejection, modulate haemostatic or thrombolytic activity, modulate
 inflammation, cancers, cardiovascular disorders, neurological disease and
 bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 the present invention.
 CC
 XX
 SQ Sequence 103 AA;
 Alignment Scores:
 Pred. No.: 8,64e-12 Length: 103
 Score: 22.00 Matches: 22
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.27% Indels: 0
 DB: 21 Gaps: 0
 US-09-966-880A-9 (1-5514) x AAB43997 (1-103)
 QY 5448 TGTATTTTGTAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTGAC 5507
 |||||
 Db 82 CysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrProAsp 101
 |||||
 QY 5508 CTCAGA 5513
 |||||
 Db 102 LeuArg 103

RESULT 4

AAB43292
 ID AAB43292 standard; Protein; 110 AA.
 XX
 AC AAB43292;
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF3056 polypeptide sequence SEQ ID NO:6112.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX Shimkets RA, Leach M;
 PI
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC77501.
 DR
 XX Novel nucleic acids and peptides derived from open reading frame X,
 useful for treating e.g. cancers, proliferative disorders,
 neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 5297-5298; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 110 AA;

Alignment Scores: 8.56e-12 Length: 110
Pred. No.: 22.00 Matches: 22
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.27% Indels: 0
DB: 21 Gaps: 0

US-09-966-880A-9 (1-5514) x AAB43292 (1-110)

QY 5448 TGTATTTTAGAGATGGGTTTCACCATGTGGCCAGGTGGTCTCAAACTCTGAC 5507
ID AAB93273 standard; Protein; 116 AA.
XX
AC AAB93273;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12312.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12312; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAB03166 to AAB13628 and
CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 116 AA;

Alignment Scores: 8.5e-12 Length: 116
Pred. No.: 22.00 Matches: 22
Score: 100.00% Conservativity: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 1.27% Gaps: 0
DB: 22

US-09-966-880A-9 (1-5514) x AAB93273 (1-116)

QY 5446 TTGTATTTTAGAGATGGGTTTCACCATGTGGCCAGGTGGTCTCAAACTCTG 5505
ID AAB93273 standard; Protein; 137 AA.
XX
AC AAB93273;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #19368.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS83564.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 49736; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 137 AA;

Alignment Scores: 8.31e-12 Length: 137
Pred. No.: 22.00 Matches: 22
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 1.27% Indels: 0
Query Match: 22 Gaps: 0
DB:

US-09-966-880A-9 (1-5514) x ABG19377 (1-137)

QY 3559 CAGGTGGTGGCTCACCGCTGTAAATCCAGCACTCTGGGAGGTGGTGGGATPCG 3618
|||||
Db 64 GlnValTrpTrpLeuThrProValIleProAlaLeuTrpGluAlaGluValGlyGlySer 83

QY 3619 CTTGAG 3624

Db 84 LeuGlu 85
|||||

RESULT 7

AAU30450
ID AAU30450 standard; Protein: 407 AA.

XX AAU30450;

XX 18-DEC-2001 (first entry)

DT Novel human secreted protein #941.

DE Human; vaccination; gene therapy; nutritional supplement;

XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.

XX WO200179449-A2.

PN 25-OCT-2001.

XX 16-APR-2001; 2001WO-US08656.

XX 18-APR-2000; 2000US-0552929.

PR 26-JAN-2001; 2001US-0770160.

XX (HYSEQ-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -

XX Claim 20; Page 294; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX Sequence 407 AA;

Alignment Scores: 7.15e-12 Length: 407
Pred. No.: 22.00 Matches: 22
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 1.27% Indels: 0
Query Match: 22 Gaps: 0
DB:

US-09-966-880A-9 (1-5514) x AAU30450 (1-407)

QY 5446 TTTGTATTTTAGAGATGGGTTTCCACCATGTTGCCAGGCTGTCTCAACTCCTG 5505
|||||
Db 63 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeuLysLeuLeu 82

QY 5506 ACCTCA 5511

Db 83 ThrSer 84
|||||

RESULT 8

AAU86859

ID AAU86859 standard; Protein: 80 AA.

XX AAU86859;

XX 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen SEQ ID NO:14452.

DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.

OS Homo sapiens.

XX WO200157182-A2.

PN 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225271.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 03-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231142.
 PR 08-SEP-2000; 2000US-0231143.
 PR 08-SEP-2000; 2000US-0231144.
 PR 08-SEP-2000; 2000US-0231145.
 PR 08-SEP-2000; 2000US-0231146.
 PR 08-SEP-2000; 2000US-0231208.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0232403.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234597.
 PR 25-SEP-2000; 2000US-0234598.
 PR 25-SEP-2000; 2000US-0234599.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

N-PSDB; AAK59640.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Claim 11; SEQ ID NO 14452; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I)

CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK34942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.

XX SQ Sequence 80 AA;

Alignment Scores: 8.24e-11 Length: 80
Pred. No.: 21.00 Matches: 21
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.21% Indels: 0
Gaps: 22

US-09-966-880A-9 (1-5514) x AAK86859 (1-80)

QY 3562 GTGTGGTCTACGCGCTGTAATCCAGCACTCTGGGAGGCTGAGTGGTGCGCTT 3621

Db 43 ValTrpIleuThrProValIleProAlaLeuTrpGluAlaGluValGlySerLeu 62
|||||

QY 3622 GAG 3624

Db 63 Glu 63
|||

RESULT 9
AAB93908

ID AAB93908 standard; Protein: 118 AA.

XX AC AAB93908;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:13871.

XX DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EPI074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isozaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

XX PS Claim 8: SEQ ID 13871; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAK03166 to AAK13628 and
CC AAK13633 to AAK18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAK13629 to AAK13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX SQ Sequence 118 AA;

Alignment Scores: 7.81e-11 Length: 118
Pred. No.: 21.00 Matches: 21
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.21% Indels: 0
Gaps: 22

US-09-966-880A-9 (1-5514) x AAB93908 (1-118)

QY 5448 TGTATTTTGTAGTAGAGATGGGTTTCACCATGTTGGCAGGCTGCTCAAACTCCTGAC 5507
|||||

Db 51 CysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrProAsp 70

QY 5508 CTC 5510

Db 71 Leu 71
|||

RESULT 10
AAO12441

ID AAO12441 standard; Protein: 36 AA.

XX AC AAO12441;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 26333.

XX DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.

XX DR N-PSDB; AAI92372.

XX PT Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune disorders -

XX Claim 20; SEQ ID NO 26333; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and the encoded proteins (AA000010-AA013910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 36 AA;

Alignment Scores:

Pred. No.:	8.47e-10	Length:	36
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.15%	Indels:	0
DB:	22	Gaps:	0

US-09-966-880A-9 (1-5514) x AA012441 (1-36)

QY 3565 TGGTGGCTCACGCTCTAATCCAGCACCTGGAGGCTGAGTGGTGGATCGCTTGAG 3624

DB 15 TrpTrpLeuThrProValIleProAlaLeuTrpGluAlaGluValGlyGlySerLeuGlu 34

RESULT 11

ABG11938

ID ABG11938 standard; Protein; 42 AA.

AC ABG11938;

XX 18-FEB-2002 (first entry)

DT DT

DE Novel human diagnostic protein #11929.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN WO200175067-A2.

XX 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS76125.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

PT Claim 20; SEQ ID No 42297; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 42 AA;

Alignment Scores:

Pred. No.:	8.3e-10	Length:	42
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.15%	Indels:	0
DB:	22	Gaps:	0

US-09-966-880A-9 (1-5514) x ABG11938 (1-42)

QY 5452 TTTTGTAGTAGATGGGTTTCACCATGTTGGCAGGCTGGTCTCAACTCTGACCTCA 5511

DB 5 PheLeuValGluMetGlyPheHisVaIGlyGlnAlaGlyLeuLysLeuThrSer 24

RESULT 12

ABG09800

ID ABG09800 standard; Protein; 89 AA.

XX AC ABG09800;

XX 13-FEB-2002 (first entry)

DT DT

DE Novel human diagnostic protein #9791.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS73987.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess

PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	(HYSE-) HYSEQ INC.
PA	
XX	
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
XX	WPI: 2001-442253/47.
DR	N-PSDB; AA161012.
DR	
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
XX	Example 2; SEQ ID NO 5787; 10078pp; English.
PS	
XX	
CC	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	the encoded polypeptides (AA38642-AA442113) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and amyotrophic lateral sclerosis. Other uses include the
CC utilisation of the activities such as: Immune system suppression,

CC activating/inhibiting activity, chemokine/chemokine receptor activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SO Sequence 125 AA: SO

Alignment Scores:

Score: NO.:	7.14E 10	Length:	223
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query match:	1.13%	0
DB:	22	0
	Gaps:	0
	Inders:	0

US-09-966-880A-9 (1-5514) x AAM41856 (1-125)

5449 GTATTTTAGTAGAGATGGGGTTTACCATGTGTGGCCAGGCTGGTCTCAAACCTCCTGACC 5508

90 valpheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeuLysLeuLeuThr 109

RESULT 14

AAO05250	
ID	AAO05250 standard; Protein; 41 AA.
XX	
AC	AAO05250;
XX	
DT	06-NOV-2001 (first entry)
XX	

DE human polypeptide SEQ ID NO 19142.
XX

KW vaccine: peptide therapy: stem cell growth factor: haematopoiesis: human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders: arthritis: inflammation

XX
05 Homo sapiens

XX
PN W0200164835-A2

XX
PD
07-SEP-2001

XX
PE
26-FEB-2001. 2001WO-US04927

PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
PI WPI: 2001-514838/56.
XX N-PSDB; AA185181.
DR Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders
XX Claim 20: SEQ ID NO 19142; 1399pp + Sequence Listing; English.
PS The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 41 AA;

Alignment Scores:
Pred. No.: 7.67e-09 Length: 41
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x AAO05250 (1-41)

QY 3614 CCACCACCTCAGCTCCAGAGTGTGGATTACAGGCTGAGCCACACACCTGG 3558
Db 21 ProProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHsthrTrp 39

RESULT 15
AAO09035
ID AAO09035 standard; Protein; 55 AA.
XX
AC AAO09035;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 22927.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX

PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
PI WPI: 2001-514838/56.
XX N-PSDB; AA188966.
DR Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders
XX Claim 20: SEQ ID NO 22927; 1399pp + Sequence Listing; English.
PS The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 55 AA;

Alignment Scores:
Pred. No.: 7.36e-09 Length: 55
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x AAO09035 (1-55)

QY 3303 GTCTTGAACCTCCTGGGCTCAAGCAATCCTCGCTGGCCTCCCAAGTGTGGGA 3359
Db 34 ValLeuAsnSerTrpAlaGlnAlaLeuLeuProTrpProPolysValleuGly 52

Search completed: June 14, 2003, 19:16:33
Job time : 234.5 secs


```
Alignment Scores:
Pred. No.: 1.22e-10 Length: 103
Score: 22.00 Matches: 22
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.27% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-925-301-1442 (1-103)
QY 5448 TGTATTTTGTAGAGATGGGTTTCACCATGTTGGCAGGCTGGTCTCAAACTCCTGCAC 5507
Db 82 CysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrProAsp 101
QY 5508 CTCAGA 5513
Db 102 LeuArg 103

RESULT 2
US-09-966-880A-9 (1-5514) x US-09-925-301-1442 (1-103)
; Sequence 4945, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4945
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (57)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (85)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (86)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (87)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (94)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (96)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (102)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (110)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (111)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
```

```
; NAME/KEY: SITE
; LOCATION: (113)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4945

Alignment Scores:
Pred. No.: 5.89e-08 Length: 115
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-891-4945 (1-115)
QY 5446 TTTTGTATTTTGTAGAGATGGGTTTCACCATGTTGGCAGGCTGGTCTCAAACTC 5502
Db 38 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeuLysLeu 56

RESULT 3
US-09-995-494-79
; Sequence 79, Application US/09995494
; Patent No. US20020127578A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Chen, Sei-Yu
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes
; FILE REFERENCE: DEX-0293
; CURRENT APPLICATION NUMBER: US/09/995,494
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/253,176
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-995-494-79

Alignment Scores:
Pred. No.: 5.89e-08 Length: 115
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-995-494-79 (1-115)
QY 3614 CCACCCACCTCAGCTCCAGAGTCTGGGATTACAGGCTGAGCCACACACCTGG 3558
Db 95 ProProThrSerAlaSerGlnSerAlaGlylleThrGlyValSerHisHisThrtrp 113

RESULT 4
US-09-764-891-4924
; Sequence 4924, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
```

; SEQ ID NO 4924
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-4924

Alignment Scores:
Pred. No.: 4.92e-07 Length: 81
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.04% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-891-4924 (1-81)

QY 3574 ACGCTGTAATCCAGCACTCTGGAGGCTGAGGTGGTGGATCGCTTGAGCCC 3627
Db 38 ThrProValIleProAlaLeuTrpGluAlaGluValGlyGlySerLeuGluPro 55
|||||

RESULT 5

US-09-764-877-1350
; Sequence 1350, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1350
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-1350

Alignment Scores:
Pred. No.: 4.07e-06 Length: 60
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-877-1350 (1-60)

QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTC 5496
Db 10 PheValPheLeuValGluMetGlyPheHisValGlyGlnAlaGlyLeu 26
|||||

RESULT 6

US-09-820-843A-111
; Sequence 111, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 111
; LENGTH: 71
; TYPE: PRT
; ORGANISM: L. major
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein P1105.01

; NAME/KEY: misc_feature
; OTHER INFORMATION: gi16996498
US-09-820-843A-111

Alignment Scores:
Pred. No.: 3.96e-06 Length: 71
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-820-843A-111 (1-71)

QY 2890 TGTGTTTGT 2840
Db 40 CysValCysValCysValCysValCysValCysValCysValCysValCysValCys 56
|||||

RESULT 7

US-09-764-891-3289
; Sequence 3289, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3289
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (82)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3289

Alignment Scores:
Pred. No.: 3.76e-06 Length: 98
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-891-3289 (1-98)

QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTC 5496
Db 16 PheValPheLeuValGluMetGlyPheHisValGlyGlnAlaGlyLeu 32
|||||

RESULT 8

US-09-908-711-115
; Sequence 115, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17

;
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 115
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (82)

;
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-115

Alignment Scores:
Pred. No.: 3.76e-06 Length: 98
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-908-711-115 (1-98)

QY 5446 TTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTC 5496
|||||
Db 16 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeu 32

RESULT 9

US-10-082-830-207
; Sequence 207, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 207
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-207

Alignment Scores:
Pred. No.: 3.61e-06 Length: 126
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-10-082-830-207 (1-126)

QY 5446 TTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTC 5496
|||||
Db 83 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeu 99

RESULT 10

US-09-925-297-833
; Sequence 833, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 2000-124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 833

; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-833

Alignment Scores:
Pred. No.: 3.79e-05 Length: 22
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-925-297-833 (1-22)

QY 5457 AGTAGAGTGGGTTTCCACCATGTGCCAGCTGGCTCAAACTCCT 5504
Db 4 SerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrPro 19
|||||

RESULT 11

US-09-925-302-742
; Sequence 742, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 742
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-742

Alignment Scores:
Pred. No.: 3.26e-05 Length: 55
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-925-302-742 (1-55)

QY 5494 GACCAGCTGCCAACATGGTGAACCCCATCTCTACTAAAATACAA 5447
Db 15 AspGlnProGlyGlnHisGlyGluThrProSerLeuLeuLysIleGln 30
|||||

RESULT 12

US-09-764-891-4886
; Sequence 4886, Application US/09764891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4886
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-4886

Alignment Scores:

Pred. No.: 3.16e-05 Length: 67
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-891-4886 (1-67)

QY 3617 GATCACCCACCTCAGCTCCAGAGTGTGGATTACAGGCGTGAGC 3570
Db 38 AspProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSer 53
|||||

RESULT 13

US-10-078-090-150
; Sequence 150, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-078-090-150

Alignment Scores:
Pred. No.: 2.99e-05 Length: 94
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-10-078-090-150 (1-94)

QY 2890 TGTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 2843
Db 35 CysValCysValCysValCysValCysValCysValCysValCysVal 50
|||||

RESULT 14

US-10-016-157A-187
; Sequence 187, Application US/10016157A
; Publication No. US20020192220A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Ghosh, Malavika
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and
; FILE REFERENCE: DEX-0253
; CURRENT APPLICATION NUMBER: US/10/016,157A
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,717
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 250
; SOFTWARE: PatentIn version 3.1

```

; SEQ ID NO 187
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-157A-187

Alignment Scores:
Pred. No.:      2.96e-05      Length:      100
Score:          16.00         Matches:      16
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:      0.92%      Indels:         0
DB:               9          Gaps:         0

US-09-966-880A-9 (1-5514) x US-10-016-157A-187 (1-100)

QY 5446 TTGTATTATTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 5493
Db 57 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGly 72

RESULT 15
US-10-074-475-280
; Sequence 280, Application US/10074475
; Publication No. US20030092898A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karia, Kalpana
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0313
; CURRENT APPLICATION NUMBER: US/10/074,475
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 60/268,292
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 295
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 280
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-074-475-280

Alignment Scores:
Pred. No.:      2.95e-05      Length:      101
Score:          16.00         Matches:      16
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:      0.92%      Indels:         0
DB:               9          Gaps:         0

US-09-966-880A-9 (1-5514) x US-10-074-475-280 (1-101)

QY 3614 CCACCCACCTCAGCTCCAGAGTGTGGGATTACAGGCTGAGCCAC 3567
Db 86 ProProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 101

Search completed: June 14, 2003, 19:43:11
Job time : 183 secs

```